short_1

November 10, 2023

1 Install neccessary packages

```
[1]: !pip install autogluon matplotlib
```

Requirement already satisfied: autogluon in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (0.8.2)
Requirement already satisfied: matplotlib in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (3.6.3)
Collecting autogluon.core[all]==0.8.2 (from autogluon)
  Obtaining dependency information for autogluon.core[all]==0.8.2 from https://f
iles.pythonhosted.org/packages/e0/56/545adb1d388e78591cd7e36de0c8b889c1944de362b
daeec0f31d01890df/autogluon.core-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.core-0.8.2-py3-none-any.whl.metadata (12 kB)
Collecting autogluon.features==0.8.2 (from autogluon)
  Obtaining dependency information for autogluon.features==0.8.2 from https://fi
les.pythonhosted.org/packages/bb/ea/7892719f78a30aee1bf42c4a0540fbae98bfbdf56b85
fab79ffc437eb687/autogluon.features-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.features-0.8.2-py3-none-any.whl.metadata (11 kB)
Collecting autogluon.tabular[all]==0.8.2 (from autogluon)
  Obtaining dependency information for autogluon.tabular[all]==0.8.2 from https:
//files.pythonhosted.org/packages/f3/dc/0bd8cadb9a5e2f3e5b12caaa6745357d912ffc7b
8b75fb4e426a38331028/autogluon.tabular-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.tabular-0.8.2-py3-none-any.whl.metadata (13 kB)
Collecting autogluon.multimodal==0.8.2 (from autogluon)
  Obtaining dependency information for autogluon.multimodal==0.8.2 from https://
files.pythonhosted.org/packages/f4/df/a3921edb866555154d5a53adbcd2268fd3b5307182
0b6382f2619ad439f3/autogluon.multimodal-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.multimodal-0.8.2-py3-none-any.whl.metadata (13 kB)
Collecting autogluon.timeseries[all]==0.8.2 (from autogluon)
  Obtaining dependency information for autogluon.timeseries[all]==0.8.2 from htt
ps://files.pythonhosted.org/packages/50/38/1768d30684292d064dad4457355d6928581c5
b832a3f65a6a023df2ff4ff/autogluon.timeseries-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.timeseries-0.8.2-py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: numpy<1.27,>=1.21 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (1.24.4)
Requirement already satisfied: scipy<1.12,>=1.5.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from

```
autogluon.core[all]==0.8.2->autogluon) (1.11.3)
Requirement already satisfied: scikit-learn<1.3,>=1.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (1.2.2)
Requirement already satisfied: networkx<4,>=3.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (3.1)
Requirement already satisfied: pandas<1.6,>=1.4.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (1.5.3)
Requirement already satisfied: tqdm<5,>=4.38 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (4.66.1)
Requirement already satisfied: requests in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (2.31.0)
Requirement already satisfied: boto3<2,>=1.10 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (1.28.60)
Collecting autogluon.common==0.8.2 (from autogluon.core[all]==0.8.2->autogluon)
  Obtaining dependency information for autogluon.common==0.8.2 from https://file
s.pythonhosted.org/packages/67/c7/aa2bd5708c9329a29245aa0cd955cace0662b92a82eb76
05cf46b6e7e9d5/autogluon.common-0.8.2-py3-none-any.whl.metadata
  Using cached autogluon.common-0.8.2-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: hyperopt<0.2.8,>=0.2.7 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (0.2.7)
Collecting ray[default]<2.4,>=2.3 (from autogluon.core[all]==0.8.2->autogluon)
  Downloading ray-2.3.1-cp310-cp310-macosx_11_0_arm64.whl (28.6 MB)

28.6/28.6 MB 8.8 MB/s eta 0:00:00m eta
0:00:01[36m0:00:01
Requirement already satisfied: pydantic<2.0,>=1.10.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.core[all]==0.8.2->autogluon) (1.10.13)
Collecting grpcio<=1.50.0,>=1.42.0 (from autogluon.core[all]==0.8.2->autogluon)
  Downloading grpcio-1.50.0.tar.gz (22.1 MB)

22.1/22.1 MB 9.9 MB/s eta 0:00:00m eta
0:00:010:01:01
  Preparing metadata (setup.py) … done
Requirement already satisfied: Pillow<9.6,>=9.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (9.5.0)
Requirement already satisfied: jsonschema<4.18,>=4.14 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (4.17.3)
Requirement already satisfied: seqeval<1.3.0,>=1.2.2 in
```

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.2.2)
Requirement already satisfied: evaluate<0.4.0,>=0.2.2 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.3.0)
Requirement already satisfied: accelerate<0.17,>=0.9 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.16.0)
Requirement already satisfied: timm<0.10.0,>=0.9.2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.9.7)
Requirement already satisfied: torch<1.14,>=1.9 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.13.1)
Requirement already satisfied: torchvision<0.15.0 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.14.1)
Requirement already satisfied: scikit-image<0.20.0,>=0.19.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.19.3)
Requirement already satisfied: pytorch-lightning<1.10.0,>=1.9.0 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.9.5)
Requirement already satisfied: text-unidecode<1.4,>=1.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.3)
Requirement already satisfied: torchmetrics<0.12.0,>=0.11.0 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.11.4)
Requirement already satisfied: transformers[sentencepiece]<4.27.0,>=4.23.0 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (4.26.1)
Requirement already satisfied: nptyping<2.5.0,>=1.4.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (2.4.1)
Requirement already satisfied: omegaconf<2.3.0,>=2.1.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (2.2.3)
Requirement already satisfied: pytorch-metric-learning<2.0,>=1.3.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.7.3)
Requirement already satisfied: nlpaug<1.2.0,>=1.1.10 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (1.1.11)
Requirement already satisfied: nltk<4.0.0,>=3.4.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (3.8.1)
Requirement already satisfied: openmim<0.4.0,>=0.3.7 in

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.3.7)
Requirement already satisfied: defusedxml<0.7.2,>=0.7.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.7.1)
Requirement already satisfied: jinja2<3.2,>=3.0.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (3.1.2)
Requirement already satisfied: tensorboard<3,>=2.9 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (2.14.1)
Requirement already satisfied: pytesseract<0.3.11,>=0.3.9 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.multimodal==0.8.2->autogluon) (0.3.10)
Requirement already satisfied: lightgbm<3.4,>=3.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.tabular[all]==0.8.2->autogluon) (3.3.5)
Requirement already satisfied: xgboost<1.8,>=1.6 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.tabular[all]==0.8.2->autogluon) (1.7.6)
Requirement already satisfied: fastai<2.8,>=2.3.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.tabular[all]==0.8.2->autogluon) (2.7.12)
Requirement already satisfied: catboost<1.3,>=1.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.tabular[all]==0.8.2->autogluon) (1.1.1)
Requirement already satisfied: joblib<2,>=1.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (1.3.2)
Requirement already satisfied: statsmodels<0.15,>=0.13.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (0.14.0)
Requirement already satisfied: gluonts<0.14,>=0.13.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (0.13.5)
Requirement already satisfied: statsforecast<1.5,>=1.4.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (1.4.0)
Requirement already satisfied: mlforecast<0.7.4,>=0.7.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (0.7.3)
Requirement already satisfied: ujson<6,>=5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.timeseries[all]==0.8.2->autogluon) (5.8.0)
Requirement already satisfied: psutil<6,>=5.7.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.common==0.8.2->autogluon.core[all]==0.8.2->autogluon) (5.9.5)
Requirement already satisfied: setuptools in

```
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
autogluon.common==0.8.2->autogluon.core[all]==0.8.2->autogluon) (68.0.0)
Requirement already satisfied: contourpy>=1.0.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(1.1.1)
Requirement already satisfied: cycler>=0.10 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(0.12.0)
Requirement already satisfied: fonttools>=4.22.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(4.43.0)
Requirement already satisfied: kiwisolver>=1.0.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(1.4.5)
Requirement already satisfied: packaging>=20.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(23.2)
Requirement already satisfied: pyparsing>=2.2.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(3.1.1)
Requirement already satisfied: python-dateutil>=2.7 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from matplotlib)
(2.8.2)
Requirement already satisfied: six>=1.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from python-
dateutil>=2.7->matplotlib) (1.16.0)
Requirement already satisfied: pyyaml in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
accelerate<0.17,>=0.9->autogluon.multimodal==0.8.2->autogluon) (6.0.1)
Requirement already satisfied: botocore<1.32.0,>=1.31.60 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
boto3<2,>=1.10->autogluon.core[all]==0.8.2->autogluon) (1.31.60)
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
boto3<2,>=1.10->autogluon.core[all]==0.8.2->autogluon) (1.0.1)
Requirement already satisfied: s3transfer<0.8.0,>=0.7.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
boto3<2,>=1.10->autogluon.core[all]==0.8.2->autogluon) (0.7.0)
Requirement already satisfied: graphviz in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
catboost<1.3,>=1.1->autogluon.tabular[all]==0.8.2->autogluon) (0.8.4)
Requirement already satisfied: plotly in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
catboost<1.3,>=1.1->autogluon.tabular[all]==0.8.2->autogluon) (5.17.0)
Requirement already satisfied: datasets>=2.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (2.14.5)
Requirement already satisfied: dill in
```

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (0.3.7)
Requirement already satisfied: xxhash in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (3.4.1)
Requirement already satisfied: multiprocess in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (0.70.15)
Requirement already satisfied: fsspec[http]>=2021.05.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (2023.6.0)
Requirement already satisfied: huggingface-hub>=0.7.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (0.17.3)
Requirement already satisfied: responses<0.19 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon) (0.18.0)
Requirement already satisfied: pip in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (23.2.1)
Requirement already satisfied: fastdownload<2,>=0.0.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (0.0.7)
Requirement already satisfied: fastcore<1.6,>=1.5.29 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (1.5.29)
Requirement already satisfied: fastprogress>=0.2.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (1.0.3)
Requirement already satisfied: spacy<4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (3.6.1)
Requirement already satisfied: toolz~=0.10 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
gluonts<0.14,>=0.13.1->autogluon.timeseries[all]==0.8.2->autogluon) (0.12.0)
Requirement already satisfied: typing-extensions~=4.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
gluonts<0.14,>=0.13.1->autogluon.timeseries[all]==0.8.2->autogluon) (4.8.0)
Requirement already satisfied: future in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==0.8.2->autogluon) (0.18.3)
Requirement already satisfied: cloudpickle in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==0.8.2->autogluon) (2.2.1)
Requirement already satisfied: py4j in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==0.8.2->autogluon) (0.10.9.7)
Requirement already satisfied: MarkupSafe>=2.0 in

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
jinja2<3.2,>=3.0.3->autogluon.multimodal==0.8.2->autogluon) (2.1.3)
Requirement already satisfied: attrs>=17.4.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
jsonschema<4.18,>=4.14->autogluon.multimodal==0.8.2->autogluon) (23.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
jsonschema<4.18,>=4.14->autogluon.multimodal==0.8.2->autogluon) (0.19.3)
Requirement already satisfied: wheel in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
lightgbm<3.4,>=3.3->autogluon.tabular[all]==0.8.2->autogluon) (0.41.2)
Requirement already satisfied: numba in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
mlforecast<0.7.4,>=0.7.0->autogluon.timeseries[all]==0.8.2->autogluon) (0.57.1)
Requirement already satisfied: window-ops in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
mlforecast<0.7.4,>=0.7.0->autogluon.timeseries[all]==0.8.2->autogluon) (0.0.14)
Requirement already satisfied: gdown>=4.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==0.8.2->autogluon) (4.7.1)
Requirement already satisfied: click in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
nltk<4.0.0,>=3.4.5->autogluon.multimodal==0.8.2->autogluon) (8.1.7)
Requirement already satisfied: regex>=2021.8.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
nltk<4.0.0,>=3.4.5->autogluon.multimodal==0.8.2->autogluon) (2023.10.3)
Requirement already satisfied: antlr4-python3-runtime==4.9.* in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
omegaconf<2.3.0,>=2.1.1->autogluon.multimodal==0.8.2->autogluon) (4.9.3)
Requirement already satisfied: colorama in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (0.4.6)
Requirement already satisfied: model-index in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (0.1.11)
Requirement already satisfied: rich in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (13.6.0)
Requirement already satisfied: tabulate in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (0.9.0)
Requirement already satisfied: pytz>=2020.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
pandas<1.6,>=1.4.1->autogluon.core[all]==0.8.2->autogluon) (2023.3.post1)
Requirement already satisfied: lightning-utilities>=0.6.0.post0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from pytorch-
lightning<1.10.0,>=1.9.0->autogluon.multimodal==0.8.2->autogluon) (0.9.0)
Requirement already satisfied: filelock in

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (3.12.4)
Requirement already satisfied: msgpack<2.0.0,>=1.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (1.0.7)
Requirement already satisfied: protobuf!=3.19.5,>=3.15.3 in
/Users/jorgensandhaug/.local/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (3.20.2)
Requirement already satisfied: aiosignal in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (1.3.1)
Requirement already satisfied: frozenlist in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (1.4.0)
Requirement already satisfied: virtualenv>=20.0.24 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (20.21.0)
  Using cached grpcio-1.49.1-cp310-cp310-macosx_11_0_arm64.whl
Requirement already satisfied: aiohttp>=3.7 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (3.8.5)
Requirement already satisfied: aiohttp-cors in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.7.0)
Requirement already satisfied: colorful in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.5.5)
Requirement already satisfied: py-spy>=0.2.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.3.14)
Requirement already satisfied: gpustat>=1.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (1.1.1)
Requirement already satisfied: opencensus in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.11.3)
Requirement already satisfied: prometheus-client>=0.7.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.17.1)
Requirement already satisfied: smart-open in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (6.4.0)
Requirement already satisfied: tensorboardX>=1.9 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (2.6.2.2)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
requests->autogluon.core[all]==0.8.2->autogluon) (3.3.0)

Requirement already satisfied: idna<4,>=2.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
requests->autogluon.core[all]==0.8.2->autogluon) (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
requests->autogluon.core[all]==0.8.2->autogluon) (1.26.17)
Requirement already satisfied: certifi>=2017.4.17 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
requests->autogluon.core[all]==0.8.2->autogluon) (2023.7.22)
Requirement already satisfied: imageio>=2.4.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from scikit-
image<0.20.0,>=0.19.1->autogluon.multimodal==0.8.2->autogluon) (2.31.1)
Requirement already satisfied: tifffile>=2019.7.26 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from scikit-
image<0.20.0,>=0.19.1->autogluon.multimodal==0.8.2->autogluon) (2023.9.26)
Requirement already satisfied: PyWavelets>=1.1.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from scikit-
image<0.20.0,>=0.19.1->autogluon.multimodal==0.8.2->autogluon) (1.4.1)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from scikit-
learn<1.3,>=1.0->autogluon.core[all]==0.8.2->autogluon) (3.2.0)
Requirement already satisfied: patsy>=0.5.2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
statsmodels<0.15,>=0.13.0->autogluon.timeseries[all]==0.8.2->autogluon) (0.5.3)
Requirement already satisfied: absl-py>=0.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (2.0.0)
Requirement already satisfied: google-auth<3,>=1.6.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (2.23.2)
Requirement already satisfied: google-auth-oauthlib<1.1,>=0.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (1.0.0)
Requirement already satisfied: markdown>=2.6.8 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (3.4.4)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (0.7.0)
Requirement already satisfied: werkzeug>=1.0.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon) (3.0.0)
Requirement already satisfied: safetensors in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
timm<0.10.0,>=0.9.2->autogluon.multimodal==0.8.2->autogluon) (0.3.3)
Requirement already satisfied: tokenizers!=0.11.3,<0.14,>=0.11.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from transformers[
sentencepiece]<4.27.0,>=4.23.0->autogluon.multimodal==0.8.2->autogluon) (0.13.3)

Requirement already satisfied: sentencepiece!=0.1.92,>=0.1.91 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from transformers[
sentencepiece]<4.27.0,>=4.23.0->autogluon.multimodal==0.8.2->autogluon) (0.1.99)
Requirement already satisfied: multidict<7.0,>=4.5 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
aiohttp>=3.7->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(6.0.4)
Requirement already satisfied: async-timeout<5.0,>=4.0.0a3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
aiohttp>=3.7->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(4.0.3)
Requirement already satisfied: yarl<2.0,>=1.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
aiohttp>=3.7->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(1.9.2)
Requirement already satisfied: pyarrow>=8.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
datasets>=2.0.0->evaluate<0.4.0,>=0.2.2->autogluon.multimodal==0.8.2->autogluon)
(13.0.0)
Requirement already satisfied: beautifulsoup4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==0.8.2->autogluon)
(4.12.2)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from google-
auth<3,>=1.6.3->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(5.3.1)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from google-
auth<3,>=1.6.3->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(0.3.0)
Requirement already satisfied: rsa<5,>=3.1.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from google-
auth<3,>=1.6.3->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(4.9)
Requirement already satisfied: requests-oauthlib>=0.7.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from google-auth-
oauthlib<1.1,>=0.5->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(1.3.1)
Requirement already satisfied: nvidia-ml-py>=11.450.129 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
gpustat>=1.0.0->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(12.535.108)
Requirement already satisfied: blessed>=1.17.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
gpustat>=1.0.0->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(1.20.0)
Requirement already satisfied: llvmlite<0.41,>=0.40.0dev0 in

/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
numba->mlforecast<0.7.4,>=0.7.0->autogluon.timeseries[all]==0.8.2->autogluon)
(0.40.1)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (1.0.10)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (2.0.8)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (3.0.9)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (8.1.12)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (1.1.2)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (2.4.8)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (2.0.10)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (0.9.0)
Requirement already satisfied: pathy>=0.10.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (0.10.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon) (3.3.0)
Requirement already satisfied: distlib<1,>=0.3.6 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from virtualenv>=2
0.0.24->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (0.3.7)
Requirement already satisfied: platformdirs<4,>=2.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from virtualenv>=2
0.0.24->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon) (3.10.0)
Requirement already satisfied: ordered-set in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from model-
index->openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (4.1.0)

```
Requirement already satisfied: opencensus-context>=0.1.3 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
opencensus->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(0.1.3)
Requirement already satisfied: google-api-core<3.0.0,>=1.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
opencensus->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2->autogluon)
(2.12.0)
Requirement already satisfied: tenacity>=6.2.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
plotly->catboost<1.3,>=1.1->autogluon.tabular[all]==0.8.2->autogluon) (8.2.3)
Requirement already satisfied: markdown-it-py>=2.2.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon) (2.16.1)
Requirement already satisfied: wcwidth>=0.1.4 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from blessed>=1.17
.1->gpustat>=1.0.0->ray[default]<2.4,>=2.3->autogluon.core[all]==0.8.2-
>autogluon) (0.2.5)
Requirement already satisfied: googleapis-common-protos<2.0.dev0,>=1.56.2 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from google-api-co
re<3.0.0,>=1.0.0->opencensus->ray[default]<2.4,>=2.3-
>autogluon.core[all]==0.8.2->autogluon) (1.60.0)
Requirement already satisfied: mdurl~=0.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from markdown-it-
py>=2.2.0->rich->openmim<0.4.0,>=0.3.7->autogluon.multimodal==0.8.2->autogluon)
(0.1.0)
Requirement already satisfied: pyasn1<0.6.0,>=0.4.6 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
pyasn1-modules>=0.2.1->google-
auth<3,>=1.6.3->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(0.5.0)
Requirement already satisfied: oauthlib>=3.0.0 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from requests-
oauthlib>=0.7.0->google-auth-
oauthlib<1.1,>=0.5->tensorboard<3,>=2.9->autogluon.multimodal==0.8.2->autogluon)
(3.2.2)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from thinc<8.2.0,>
=8.1.8->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon)
(0.7.10)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from thinc<8.2.0,>
=8.1.8->spacy<4->fastai<2.8,>=2.3.1->autogluon.tabular[all]==0.8.2->autogluon)
(0.1.3)
Requirement already satisfied: soupsieve>1.2 in
```

```
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from beautifulsoup
4->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==0.8.2->autogluon)
(2.5)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in
/opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages (from
requests->autogluon.core[all]==0.8.2->autogluon) (1.7.1)
Downloading autogluon.features-0.8.2-py3-none-any.whl (62 kB)

62.1/62.1 kB 5.7 MB/s eta 0:00:00
Downloading autogluon.multimodal-0.8.2-py3-none-any.whl (372 kB)

372.3/372.3 kB 9.4 MB/s eta 0:00:000m eta
0:00:01
Downloading autogluon.common-0.8.2-py3-none-any.whl (61 kB)

61.1/61.1 kB 5.6 MB/s eta 0:00:00
Downloading autogluon.core-0.8.2-py3-none-any.whl (224 kB)

224.0/224.0 kB 2.3 MB/s eta 0:00:00 MB/s eta
0:00:01:01
Downloading autogluon.tabular-0.8.2-py3-none-any.whl (285 kB)

285.7/285.7 kB 9.4 MB/s eta 0:00:00
Downloading autogluon.timeseries-0.8.2-py3-none-any.whl (116 kB)

116.3/116.3 kB 7.7 MB/s eta 0:00:00
Installing collected packages: grpcio, ray, autogluon.common,
autogluon.features, autogluon.core, autogluon.tabular, autogluon.multimodal,
autogluon.timeseries
  Attempting uninstall: ray
    Found existing installation: ray 2.6.3
    Uninstalling ray-2.6.3:
      Successfully uninstalled ray-2.6.3
  Attempting uninstall: autogluon.common
    Found existing installation: autogluon.common 0.8.1
    Uninstalling autogluon.common-0.8.1:
^C
ERROR: Operation cancelled by user
```

2 Config

```python
[45]:  # config

       label = 'y'
       metric = 'mean_absolute_error'
```

```
time_limit = None
presets = "experimental_zeroshot_hpo_hybrid"#'best_quality'

use_is_estimated_attr = True

drop_night_outliers = True

# to_drop = ["snow_drift:idx", "snow_density:kgm3", "wind_speed_w_1000hPa:ms",
# ↪"dew_or_rime:idx", "prob_rime:p", "fresh_snow_12h:cm", "fresh_snow_24h:cm",
# ↪"wind_speed_u_10m:ms", "wind_speed_v_10m:ms", "snow_melt_10min:mm",
# ↪"rain_water:kgm2", "dew_point_2m:K", "precip_5min:mm", "absolute_humidity_2m:
# ↪gm3", "air_density_2m:kgm3"]#, "msl_pressure:hPa", "pressure_50m:hPa",
# ↪"pressure_100m:hPa"]
to_drop = []#["wind_speed_w_1000hPa:ms", "wind_speed_u_10m:ms",
# ↪"wind_speed_v_10m:ms"]

excluded_model_types = ['CAT', 'XGB', 'RF']

num_stack_levels = 0
num_bag_folds = 0#8# 8
num_bag_sets = 0##20

use_tune_data = True
use_test_data = True
use_bag_holdout = True

clip_predictions = True
```

3 Loading and preprocessing

```
[46]: import pandas as pd
      import numpy as np



      import warnings
      warnings.filterwarnings("ignore")


      def feature_engineering(X):
          # shift columns with them by 1 hour, so that for index 16:00, we have the
      ↪values from 17:00
          columns = ['clear_sky_energy_1h:J', 'diffuse_rad_1h:J', 'direct_rad_1h:J',
                     'fresh_snow_12h:cm', 'fresh_snow_1h:cm', 'fresh_snow_24h:cm',
                     'fresh_snow_3h:cm', 'fresh_snow_6h:cm']
```

```python
    # Filter rows where index.minute == 0
    X_shifted = X[X.index.minute == 0][columns].copy()

    # Create a set for constant-time lookup
    index_set = set(X.index)

    # Vectorized time shifting
    one_hour = pd.Timedelta('1 hour')
    shifted_indices = X_shifted.index + one_hour
    X_shifted.loc[shifted_indices.isin(index_set)] = X.
↪loc[shifted_indices[shifted_indices.isin(index_set)]][columns]

    # set last row to same as second last row
    X_shifted.iloc[-1] = X_shifted.iloc[-2]


    # Rename columns
    X_old_unshifted = X_shifted.copy()
    X_old_unshifted.columns = [f"{col}_not_shifted" for col in X_old_unshifted.
↪columns]

    date_calc = None
    # If 'date_calc' is present, handle it
    if 'date_calc' in X.columns:
        date_calc = X[X.index.minute == 0]['date_calc']



    # resample to hourly
    X = X.resample('H').mean()

    # overwrite columns with shifted columns
    X[columns] = X_shifted[columns]

    if date_calc is not None:
        X['date_calc'] = date_calc

    return X



def fix_X(X, name):
    # Convert 'date_forecast' to datetime format and replace original column␣
↪with 'ds'
    X['ds'] = pd.to_datetime(X['date_forecast'])
```

```python
    X.drop(columns=['date_forecast'], inplace=True, errors='ignore')
    X.sort_values(by='ds', inplace=True)
    X.set_index('ds', inplace=True)


    X = feature_engineering(X)

    return X



def handle_features(X_train_observed, X_train_estimated, X_test, y_train):
    X_train_observed = fix_X(X_train_observed, "X_train_observed")
    X_train_estimated = fix_X(X_train_estimated, "X_train_estimated")
    X_test = fix_X(X_test, "X_test")


    y_train['ds'] = pd.to_datetime(y_train['time'])
    y_train.drop(columns=['time'], inplace=True)
    y_train.sort_values(by='ds', inplace=True)
    y_train.set_index('ds', inplace=True)

    return X_train_observed, X_train_estimated, X_test, y_train



def preprocess_data(X_train_observed, X_train_estimated, X_test, y_train,␣
 ↪location):
    # convert to datetime
    X_train_observed, X_train_estimated, X_test, y_train =␣
 ↪handle_features(X_train_observed, X_train_estimated, X_test, y_train)

    if use_is_estimated_attr:
        X_train_observed["is_estimated"] = 0
        X_train_estimated["is_estimated"] = 1
        X_test["is_estimated"] = 1

    # drop date_calc
    X_train_estimated.drop(columns=['date_calc'], inplace=True)
    X_test.drop(columns=['date_calc'], inplace=True)


    y_train["y"] = y_train["pv_measurement"].astype('float64')
    y_train.drop(columns=['pv_measurement'], inplace=True)
    X_train = pd.concat([X_train_observed, X_train_estimated])
```

```python
    # clip all y values to 0 if negative
    y_train["y"] = y_train["y"].clip(lower=0)

    X_train = pd.merge(X_train, y_train, how="inner", left_index=True,␣
 ↪right_index=True)


    X_train["location"] = location
    X_test["location"] = location

    return X_train, X_test

# Define locations
locations = ['A', 'B', 'C']

X_trains = []
X_tests = []
# Loop through locations
for loc in locations:
    print(f"Processing location {loc}...")
    # Read target training data
    y_train = pd.read_parquet(f'{loc}/train_targets.parquet')

    # Read estimated training data and add location feature
    X_train_estimated = pd.read_parquet(f'{loc}/X_train_estimated.parquet')

    # Read observed training data and add location feature
    X_train_observed= pd.read_parquet(f'{loc}/X_train_observed.parquet')

    # Read estimated test data and add location feature
    X_test_estimated = pd.read_parquet(f'{loc}/X_test_estimated.parquet')

    # Preprocess data
    X_train, X_test = preprocess_data(X_train_observed, X_train_estimated,␣
 ↪X_test_estimated, y_train, loc)

    X_trains.append(X_train)
    X_tests.append(X_test)

# Concatenate all data and save to csv
X_train = pd.concat(X_trains)
X_test = pd.concat(X_tests)
```

Processing location A…
Processing location B…
Processing location C…

3.1 Feature enginering

3.1.1 Remove anomalies

```python
[47]: import pandas as pd

def replace_streaks_with_nan(df, max_streak_length, column="y"):
    for location in df["location"].unique():
        x = df[df["location"] == location][column].copy()

        last_val = None
        streak_length = 1
        streak_indices = []
        allowed = [0]
        found_streaks = {}

        for idx in x.index:
            value = x[idx]

            if value == last_val and value not in allowed:
                streak_length += 1
                streak_indices.append(idx)
            else:
                streak_length = 1
                last_val = value
                streak_indices.clear()

            if streak_length > max_streak_length:
                found_streaks[value] = streak_length

                for streak_idx in streak_indices:
                    x[idx] = np.nan
                streak_indices.clear()  # clear after setting to NaN to avoid␣
    ↪setting multiple times
        df.loc[df["location"] == location, column] = x

        print(f"Found streaks for location {location}: {found_streaks}")

    return df


X_train = replace_streaks_with_nan(X_train.copy(), 3, "y")
```

Found streaks for location A: {}
Found streaks for location B: {3.45: 28, 6.9: 7, 12.9375: 5, 13.8: 8, 276.0: 78,
18.975: 58, 0.8625: 4, 118.1625: 33, 34.5: 11, 183.7125: 1058, 87.1125: 7,
79.35: 34, 7.7625: 12, 27.6: 448, 273.41249999999997: 72, 264.78749999999997:
55, 169.05: 33, 375.1875: 56, 314.8125: 66, 76.7625: 10, 135.4125: 216, 81.9375:
202, 2.5875: 12, 81.075: 210}

18

```
Found streaks for location C: {9.8: 4, 29.400000000000002: 4, 19.6: 4}
```

```python
[48]: # print num rows
      temprows = len(X_train)
      X_train.dropna(subset=['y', 'direct_rad_1h:J', 'diffuse_rad_1h:J'],␣
       ↪inplace=True)
      print("Dropped rows: ", temprows - len(X_train))
```

```
Dropped rows:  9293
```

```python
[49]: thresh = 0.1
      mask = (X_train["direct_rad_1h:J"] <= thresh) & (X_train["diffuse_rad_1h:J"] <=␣
       ↪thresh) & (X_train["y"] >= 0.1)

      if drop_night_outliers:
          X_train.loc[mask, "y"] = np.nan
```

```python
[50]: temprows = len(X_train)
      X_train.dropna(subset=['y', 'direct_rad_1h:J', 'diffuse_rad_1h:J'],␣
       ↪inplace=True)
      print("Dropped rows: ", temprows - len(X_train))
```

```
Dropped rows:  1876
```

```python
[51]: X_train.drop(columns=to_drop, inplace=True)
      X_test.drop(columns=to_drop, inplace=True)

      X_train.to_csv('X_train_raw.csv', index=True)
      X_test.to_csv('X_test_raw.csv', index=True)
```

```python
[52]: def split_and_shuffle_data(input_data, num_bins, frac1):
          # Validate the input fraction
          if frac1 < 0 or frac1 > 1:
              raise ValueError("frac1 must be between 0 and 1.")

          if frac1==1:
              return input_data, pd.DataFrame()

          # Calculate the fraction for the second output set
          frac2 = 1 - frac1

          # Calculate bin size
          bin_size = len(input_data) // num_bins

          # Initialize empty DataFrames for output
          output_data1 = pd.DataFrame()
          output_data2 = pd.DataFrame()
```

```python
    for i in range(num_bins):
        # Shuffle the data in the current bin
        np.random.seed(i)
        current_bin = input_data.iloc[i * bin_size: (i + 1) * bin_size].
    ↪sample(frac=1)

        # Calculate the sizes for each output set
        size1 = int(len(current_bin) * frac1)

        # Split and append to output DataFrames
        output_data1 = pd.concat([output_data1, current_bin.iloc[:size1]])
        output_data2 = pd.concat([output_data2, current_bin.iloc[size1:]])

    # Shuffle and split the remaining data
    remaining_data = input_data.iloc[num_bins * bin_size:].sample(frac=1)

    remaining_size1 = int(len(remaining_data) * frac1)

    output_data1 = pd.concat([output_data1, remaining_data.iloc[:
    ↪remaining_size1]])
    output_data2 = pd.concat([output_data2, remaining_data.iloc[remaining_size1:
    ↪]])

    return output_data1, output_data2
```

```python
[53]: from autogluon.tabular import TabularDataset, TabularPredictor
      data = TabularDataset('X_train_raw.csv')
      data['ds'] = pd.to_datetime(data['ds'])
      data = data.sort_values(by='ds')

      split_time = pd.to_datetime("2022-10-28 22:00:00")
      train_set = TabularDataset(data[data["ds"] < split_time])
      estimated_set = TabularDataset(data[data["ds"] >= split_time]) # only estimated

      test_set = pd.DataFrame()
      tune_set = pd.DataFrame()
      new_train_set = pd.DataFrame()

      for location in locations:
          loc_data = data[data["location"] == location]
          num_train_rows = len(loc_data)

          tune_rows = 1500.0 # 2500.0
          if use_test_data:
              tune_rows = 1880.0#max(3000.0,␣
      ↪len(estimated_set[estimated_set["location"] == location]))
```

```python
        holdout_frac = max(0.01, min(0.1, tune_rows / num_train_rows)) *␣
↪num_train_rows / len(estimated_set[estimated_set["location"] == location])

        # shuffle and split data
        loc_tune_set, loc_new_train_set =␣
↪split_and_shuffle_data(estimated_set[estimated_set['location'] == location],␣
↪40, holdout_frac)

    new_train_set = pd.concat([new_train_set, loc_new_train_set])

    if use_test_data:
        loc_test_set, loc_tune_set = split_and_shuffle_data(loc_tune_set, 40, 0.
↪2)
        test_set = pd.concat([test_set, loc_test_set])

    tune_set = pd.concat([tune_set, loc_tune_set])

# add rest to train_set
train_set = pd.concat([train_set, new_train_set])
tuning_data = tune_set
if use_test_data:
    test_data = test_set

train_data = train_set

train_data = TabularDataset(train_data)
tuning_data = TabularDataset(tuning_data)

if use_test_data:
    test_data = TabularDataset(test_data)
```

Loaded data from: X_train_raw.csv | Columns = 49 / 49 | Rows = 87917 -> 87917

4 Modeling

```python
[54]: import os


# if submissions folder does not exist, create it
if not os.path.exists('submissions'):
    os.makedirs('submissions')

# Get the last submission number
last_submission_number = int(max([int(filename.split('_')[1].split('.')[0]) for␣
↪filename in os.listdir('submissions') if "submission" in filename]))
print("Last submission number:", last_submission_number)
```

```python
print("Now creating submission number:", last_submission_number + 1)

# Create the new filename
new_filename = f'submission_{last_submission_number + 1}'

print("New filename:", new_filename)
```

```
Last submission number: 133
Now creating submission number: 134
New filename: submission_134
```

[55]:
```python
predictors = [None, None, None]
```

[56]:
```python
def fit_predictor_for_location(loc):
    # All of these hyperparameters have been found by experimenting with some
    ↪standard parameters in AutoGluon, and then only using the best ones for each
    ↪location to make the train time shorter
    r118 = {'extra_trees': True, 'feature_fraction': 0.7832570544199176,
    ↪'learning_rate': 0.021720607471727896, 'min_data_in_leaf': 3, 'num_leaves':
    ↪21, 'ag_args': {'name_suffix': '_r118', 'priority': 17}}
    r51 = {'bs': 1024, 'emb_drop': 0.6046989241462619, 'epochs': 48, 'layers':
    ↪[200, 100, 50], 'lr': 0.0077530942164966, 'ps': 0.09244767444160731,
    ↪'ag_args': {'name_suffix': '_r51', 'priority': 12}}
    r145 = {'bs': 128, 'emb_drop': 0.44339037504795686, 'epochs': 31, 'layers':
    ↪[400, 200, 100], 'lr': 0.008615195908919904, 'ps': 0.19220253419114286,
    ↪'ag_args': {'name_suffix': '_r145', 'priority': 9}}
    lgbmXT = {'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}
    if loc == "A":
        hyperparameters = {
            'NN_TORCH': {},
            'GBM': [lgbmXT, r118],#, 'GBMLarge'],#, r_118],
            'FASTAI': [r51],
        }
    elif loc == "B":
        hyperparameters = {
            'NN_TORCH': {},
            'XT': [{'criterion': 'squared_error', 'ag_args': {'name_suffix':
    ↪'MSE', 'problem_types': ['regression', 'quantile']}}],
            'GBM': [lgbmXT, r118],
            'FASTAI': [{}, r51, r145],
        }

    elif loc == "C":
        hyperparameters = {
            'NN_TORCH': {},
                'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix':
    ↪'Unif'}}],
```

```python
            'GBM': [lgbmXT, r118],
            'FASTAI': [r51],
        }


    predictor = TabularPredictor(
        label=label,
        eval_metric=metric,
        path=f"AutogluonModels/{new_filename}_{loc}",
    ).fit(
        train_data=train_data[train_data["location"] == loc].
↪drop(columns=["ds"]),
        time_limit=time_limit,
        presets=presets,
        num_stack_levels=num_stack_levels,
        num_bag_folds=num_bag_folds,
        num_bag_sets=num_bag_sets,
        tuning_data=tuning_data[tuning_data["location"] == loc].
↪reset_index(drop=True).drop(columns=["ds"]),
        hyperparameters = hyperparameters,
        use_bag_holdout=use_bag_holdout,
        #excluded_model_types=excluded_model_types
    )

    # evaluate on test data
    if use_test_data:
        t = test_data[test_data["location"] == loc]
        perf = predictor.evaluate(t)
        print("Evaluation on test data:")
        print(perf[predictor.eval_metric.name])

    return predictor

loc = "A"
predictors[0] = fit_predictor_for_location(loc)
```

Warning: path already exists! This predictor may overwrite an existing
predictor! path="AutogluonModels/submission_134_A"
Presets specified: ['experimental_zeroshot_hpo_hybrid']
Stack configuration (auto_stack=True): num_stack_levels=0, num_bag_folds=0,
num_bag_sets=0
Beginning AutoGluon training …
AutoGluon will save models to "AutogluonModels/submission_134_A/"
AutoGluon Version: 0.8.2
Python Version: 3.10.12
Operating System: Darwin
Platform Machine: arm64

```
Platform Version:   Darwin Kernel Version 22.1.0: Sun Oct  9 20:15:09 PDT 2022;
root:xnu-8792.41.9~2/RELEASE_ARM64_T6000
Disk Space Avail:   101.27 GB / 494.38 GB (20.5%)
Train Data Rows:    30934
Train Data Columns: 47
Tuning Data Rows:    1485
Tuning Data Columns: 47
Label Column: y
Preprocessing data …
AutoGluon infers your prediction problem is: 'regression' (because dtype of
label-column == float and many unique label-values observed).
        Label info (max, min, mean, stddev): (5733.42, 0.0, 673.41535, 1195.24)
        If 'regression' is not the correct problem_type, please manually specify
the problem_type parameter during predictor init (You may specify problem_type
as one of: ['binary', 'multiclass', 'regression'])
Using Feature Generators to preprocess the data …
Fitting AutoMLPipelineFeatureGenerator…
        Available Memory:                      3225.39 MB
        Train Data (Original)  Memory Usage: 13.81 MB (0.4% of available memory)
        Inferring data type of each feature based on column values. Set
feature_metadata_in to manually specify special dtypes of the features.
        Stage 1 Generators:
                Fitting AsTypeFeatureGenerator…
                        Note: Converting 2 features to boolean dtype as they
only contain 2 unique values.
        Stage 2 Generators:
                Fitting FillNaFeatureGenerator…
        Stage 3 Generators:
                Fitting IdentityFeatureGenerator…
        Stage 4 Generators:
                Fitting DropUniqueFeatureGenerator…
        Stage 5 Generators:
                Fitting DropDuplicatesFeatureGenerator…
        Useless Original Features (Count: 3): ['elevation:m', 'snow_drift:idx',
'location']
                These features carry no predictive signal and should be manually
investigated.
                This is typically a feature which has the same value for all
rows.
                These features do not need to be present at inference time.
        Types of features in original data (raw dtype, special dtypes):
                ('float', []) : 43 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', [])   :  1 | ['is_estimated']
        Types of features in processed data (raw dtype, special dtypes):
                ('float', [])     : 42 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
```

```
'clear_sky_rad:W', …]
                  ('int', ['bool']) :  2 | ['snow_density:kgm3', 'is_estimated']
        0.1s = Fit runtime
        44 features in original data used to generate 44 features in processed
data.
        Train Data (Processed) Memory Usage: 10.96 MB (0.3% of available memory)
Data preprocessing and feature engineering runtime = 0.18s …
AutoGluon will gauge predictive performance using evaluation metric:
'mean_absolute_error'
        This metric's sign has been flipped to adhere to being higher_is_better.
The metric score can be multiplied by -1 to get the metric value.
        To change this, specify the eval_metric parameter of Predictor()
Warning: use_bag_holdout=True, but bagged mode is not enabled. use_bag_holdout
will be ignored.
User-specified model hyperparameters to be fit:
{
        'NN_TORCH': {},
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}],
}
Fitting 2 L1 models …
Fitting model: LightGBMXT …

[1000]  valid_set's l1: 94.1472
[2000]  valid_set's l1: 90.6203
[3000]  valid_set's l1: 89.7064
[4000]  valid_set's l1: 88.5809
[5000]  valid_set's l1: 88.1969
[6000]  valid_set's l1: 87.7129
[7000]  valid_set's l1: 87.4664
[8000]  valid_set's l1: 87.4638
[9000]  valid_set's l1: 87.3975
[10000] valid_set's l1: 87.2144

        -87.2056         = Validation score    (-mean_absolute_error)
        59.56s   = Training    runtime
        0.55s    = Validation runtime
Fitting model: NeuralNetTorch …
        -88.8725         = Validation score    (-mean_absolute_error)
        28.64s   = Training    runtime
        0.01s    = Validation runtime
Fitting model: WeightedEnsemble_L2 …
        -82.5454         = Validation score    (-mean_absolute_error)
        0.03s    = Training    runtime
        0.0s     = Validation runtime
AutoGluon training complete, total runtime = 89.45s … Best model:
"WeightedEnsemble_L2"
TabularPredictor saved. To load, use: predictor =
TabularPredictor.load("AutogluonModels/submission_134_A/")
Evaluation: mean_absolute_error on test data: -108.09115440484666
```

```
        Note: Scores are always higher_is_better. This metric score can be
multiplied by -1 to get the metric value.
Evaluations on test data:
{
    "mean_absolute_error": -108.09115440484666,
    "root_mean_squared_error": -328.46462204134156,
    "mean_squared_error": -107889.00793276135,
    "r2": 0.8306773015067175,
    "pearsonr": 0.9171184125361762,
    "median_absolute_error": -5.514040470123291
}

Evaluation on test data:
-108.09115440484666
```

```python
import matplotlib.pyplot as plt
leaderboards = [None, None, None]
def leaderboard_for_location(i, loc):
    plt.scatter(train_data[(train_data["location"] == loc) &
 ↪(train_data["is_estimated"]==True)]["y"].index,
 ↪train_data[(train_data["location"] == loc) &
 ↪(train_data["is_estimated"]==True)]["y"])
    plt.scatter(tuning_data[tuning_data["location"] == loc]["y"].index,
 ↪tuning_data[tuning_data["location"] == loc]["y"])
    plt.title("Val and Train")
    plt.show()

    if use_test_data:
        lb = predictors[i].leaderboard(test_data[test_data["location"] == loc])
        lb["location"] = loc
        plt.scatter(test_data[test_data["location"] == loc]["y"].index,
 ↪test_data[test_data["location"] == loc]["y"])
        plt.title("Test")

        return lb

    return pd.DataFrame()

leaderboards[0] = leaderboard_for_location(0, loc)
```
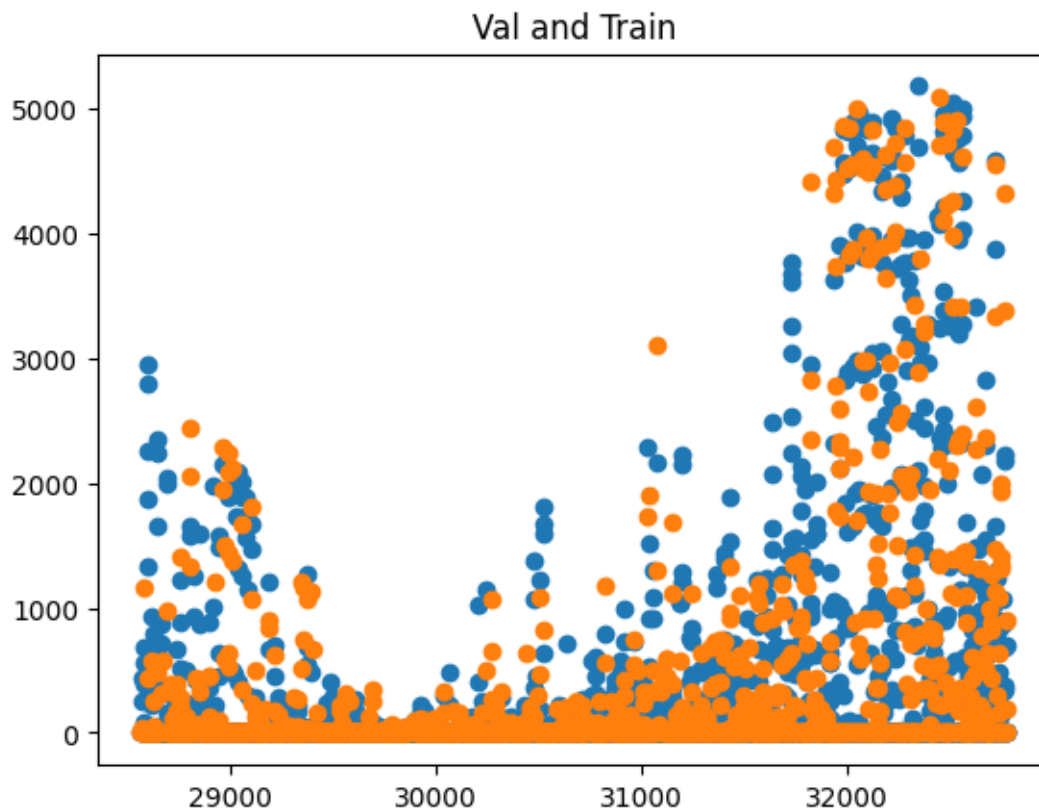
```
              model   score_test   score_val   pred_time_test   pred_time_val
fit_time  pred_time_test_marginal   pred_time_val_marginal   fit_time_marginal
stack_level   can_infer   fit_order
0          LightGBMXT -105.076269  -87.205583         0.127806        0.554786
59.564247                 0.127806                 0.554786         59.564247
1       True           1
1  WeightedEnsemble_L2 -108.091154  -82.545415         0.144621        0.569622
88.234285                 0.001212                 0.000299          0.027779
2       True           3
2      NeuralNetTorch -121.957260  -88.872490         0.015603        0.014537
28.642259                 0.015603                 0.014537         28.642259
1       True           2
```

Test
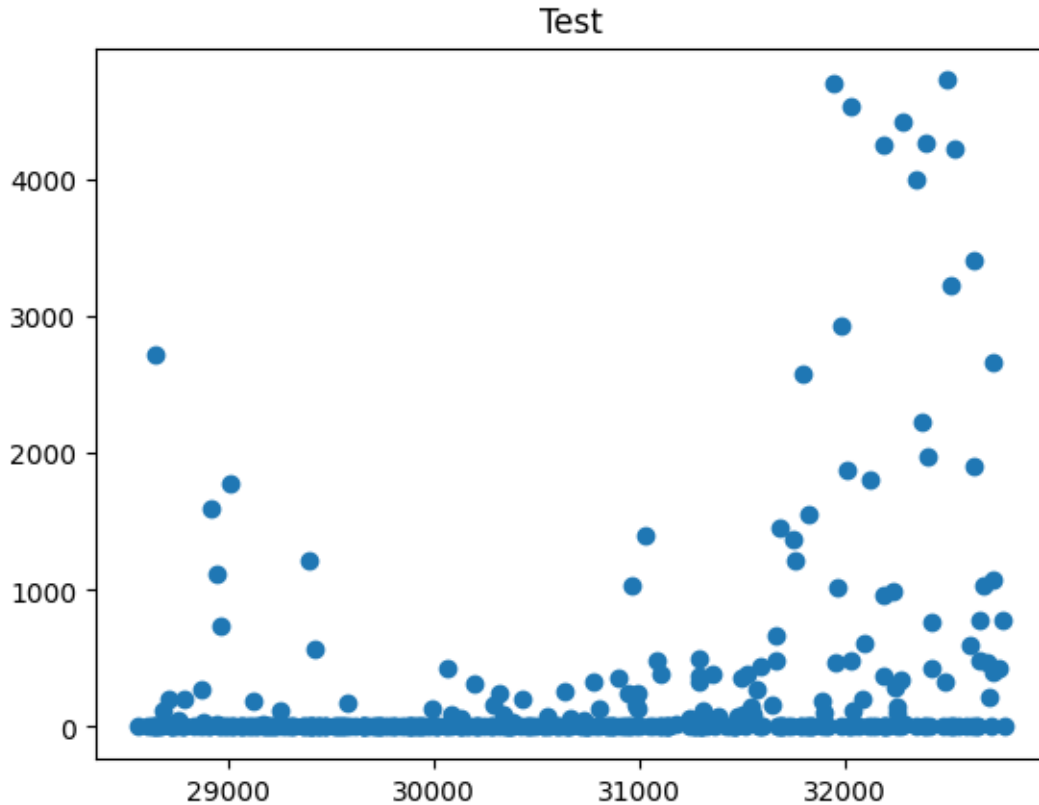
```
[58]: loc = "B"
      predictors[1] = fit_predictor_for_location(loc)
      leaderboards[1] = leaderboard_for_location(1, loc)
```

Warning: path already exists! This predictor may overwrite an existing
predictor! path="AutogluonModels/submission_134_B"
Presets specified: ['experimental_zeroshot_hpo_hybrid']
Stack configuration (auto_stack=True): num_stack_levels=0, num_bag_folds=0,
num_bag_sets=0
Beginning AutoGluon training …
AutoGluon will save models to "AutogluonModels/submission_134_B/"
AutoGluon Version: 0.8.2
Python Version: 3.10.12
Operating System: Darwin
Platform Machine: arm64
Platform Version: Darwin Kernel Version 22.1.0: Sun Oct 9 20:15:09 PDT 2022;
root:xnu-8792.41.9~2/RELEASE_ARM64_T6000
Disk Space Avail: 101.00 GB / 494.38 GB (20.4%)
Train Data Rows: 27377
Train Data Columns: 47
Tuning Data Rows: 1485

28

```
Tuning Data Columns: 47
Label Column: y
Preprocessing data …
AutoGluon infers your prediction problem is: 'regression' (because dtype of
label-column == float and many unique label-values observed).
        Label info (max, min, mean, stddev): (1152.3, -0.0, 98.11625, 206.48535)
        If 'regression' is not the correct problem_type, please manually specify
the problem_type parameter during predictor init (You may specify problem_type
as one of: ['binary', 'multiclass', 'regression'])
Using Feature Generators to preprocess the data …
Fitting AutoMLPipelineFeatureGenerator…
        Available Memory:                       2794.4 MB
        Train Data (Original)  Memory Usage: 12.3 MB (0.4% of available memory)
        Inferring data type of each feature based on column values. Set
feature_metadata_in to manually specify special dtypes of the features.
        Stage 1 Generators:
                Fitting AsTypeFeatureGenerator…
                        Note: Converting 2 features to boolean dtype as they
only contain 2 unique values.
        Stage 2 Generators:
                Fitting FillNaFeatureGenerator…
        Stage 3 Generators:
                Fitting IdentityFeatureGenerator…
        Stage 4 Generators:
                Fitting DropUniqueFeatureGenerator…
        Stage 5 Generators:
                Fitting DropDuplicatesFeatureGenerator…
        Useless Original Features (Count: 2): ['elevation:m', 'location']
                These features carry no predictive signal and should be manually
investigated.
                This is typically a feature which has the same value for all
rows.
                These features do not need to be present at inference time.
        Types of features in original data (raw dtype, special dtypes):
                ('float', []) : 44 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', [])   :  1 | ['is_estimated']
        Types of features in processed data (raw dtype, special dtypes):
                ('float', [])    : 43 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', ['bool']) :  2 | ['snow_density:kgm3', 'is_estimated']
        0.1s = Fit runtime
        45 features in original data used to generate 45 features in processed
data.
        Train Data (Processed) Memory Usage: 9.99 MB (0.4% of available memory)
Data preprocessing and feature engineering runtime = 0.14s …
```

AutoGluon will gauge predictive performance using evaluation metric:
'mean_absolute_error'
        This metric's sign has been flipped to adhere to being higher_is_better.
The metric score can be multiplied by -1 to get the metric value.
        To change this, specify the eval_metric parameter of Predictor()
Warning: use_bag_holdout=True, but bagged mode is not enabled. use_bag_holdout
will be ignored.
User-specified model hyperparameters to be fit:
{
        'NN_TORCH': {},
        'XT': [{'criterion': 'squared_error', 'ag_args': {'name_suffix': 'MSE',
'problem_types': ['regression', 'quantile']}}],
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}},
{'extra_trees': True, 'feature_fraction': 0.7832570544199176, 'learning_rate':
0.021720607471727896, 'min_data_in_leaf': 3, 'num_leaves': 21, 'ag_args':
{'name_suffix': '_r118', 'priority': 17}}],
        'FASTAI': [{}, {'bs': 1024, 'emb_drop': 0.6046989241462619, 'epochs':
48, 'layers': [200, 100, 50], 'lr': 0.00775309042164966, 'ps':
0.09244767444160731, 'ag_args': {'name_suffix': '_r51', 'priority': 12}}, {'bs':
128, 'emb_drop': 0.44339037504795686, 'epochs': 31, 'layers': [400, 200, 100],
'lr': 0.008615195908919904, 'ps': 0.19220253419114286, 'ag_args':
{'name_suffix': '_r145', 'priority': 9}}],
}
Fitting 7 L1 models …
Fitting model: LightGBMXT …

[1000] valid_set's l1: 14.2773
[2000] valid_set's l1: 14.0181
[3000] valid_set's l1: 13.957
[4000] valid_set's l1: 13.8745
[5000] valid_set's l1: 13.8621
[6000] valid_set's l1: 13.8212
[7000] valid_set's l1: 13.8141
[8000] valid_set's l1: 13.7824
[9000] valid_set's l1: 13.7291
[10000] valid_set's l1: 13.7163

        -13.7142   = Validation score   (-mean_absolute_error)
        126.65s = Training   runtime
        0.26s   = Validation runtime
Fitting model: ExtraTreesMSE …
        -15.4761   = Validation score   (-mean_absolute_error)
        3.82s   = Training   runtime
        0.05s   = Validation runtime
Fitting model: NeuralNetFastAI …
        -14.9623   = Validation score   (-mean_absolute_error)
        17.81s   = Training   runtime
        0.02s   = Validation runtime
Fitting model: NeuralNetTorch …

```
        -13.3344          = Validation score   (-mean_absolute_error)
        20.55s   = Training    runtime
        0.01s    = Validation runtime
Fitting model: LightGBM_r118 …

[1000]  valid_set's l1: 13.9149
[2000]  valid_set's l1: 13.4226
[3000]  valid_set's l1: 13.2471
[4000]  valid_set's l1: 13.1382
[5000]  valid_set's l1: 13.0259
[6000]  valid_set's l1: 12.9035
[7000]  valid_set's l1: 12.7973
[8000]  valid_set's l1: 12.7556
[9000]  valid_set's l1: 12.7006
[10000] valid_set's l1: 12.658

        -12.6552          = Validation score   (-mean_absolute_error)
        43.66s   = Training    runtime
        0.15s    = Validation runtime
Fitting model: NeuralNetFastAI_r51 …
        -14.1868          = Validation score   (-mean_absolute_error)
        11.89s   = Training    runtime
        0.01s    = Validation runtime
Fitting model: NeuralNetFastAI_r145 …
        -13.7911          = Validation score   (-mean_absolute_error)
        38.66s   = Training    runtime
        0.02s    = Validation runtime
Fitting model: WeightedEnsemble_L2 …
        -11.9643          = Validation score   (-mean_absolute_error)
        0.07s    = Training    runtime
        0.0s     = Validation runtime
AutoGluon training complete, total runtime = 264.88s … Best model:
"WeightedEnsemble_L2"
TabularPredictor saved. To load, use: predictor =
TabularPredictor.load("AutogluonModels/submission_134_B/")
Evaluation: mean_absolute_error on test data: -10.771986388368392
        Note: Scores are always higher_is_better. This metric score can be
multiplied by -1 to get the metric value.
Evaluations on test data:
{
    "mean_absolute_error": -10.771986388368392,
    "root_mean_squared_error": -29.819237777904075,
    "mean_squared_error": -889.1869416551816,
    "r2": 0.9617196394878705,
    "pearsonr": 0.9806798548041709,
    "median_absolute_error": -0.6749260425567627
}

Evaluation on test data:
```

-10.771986388368392
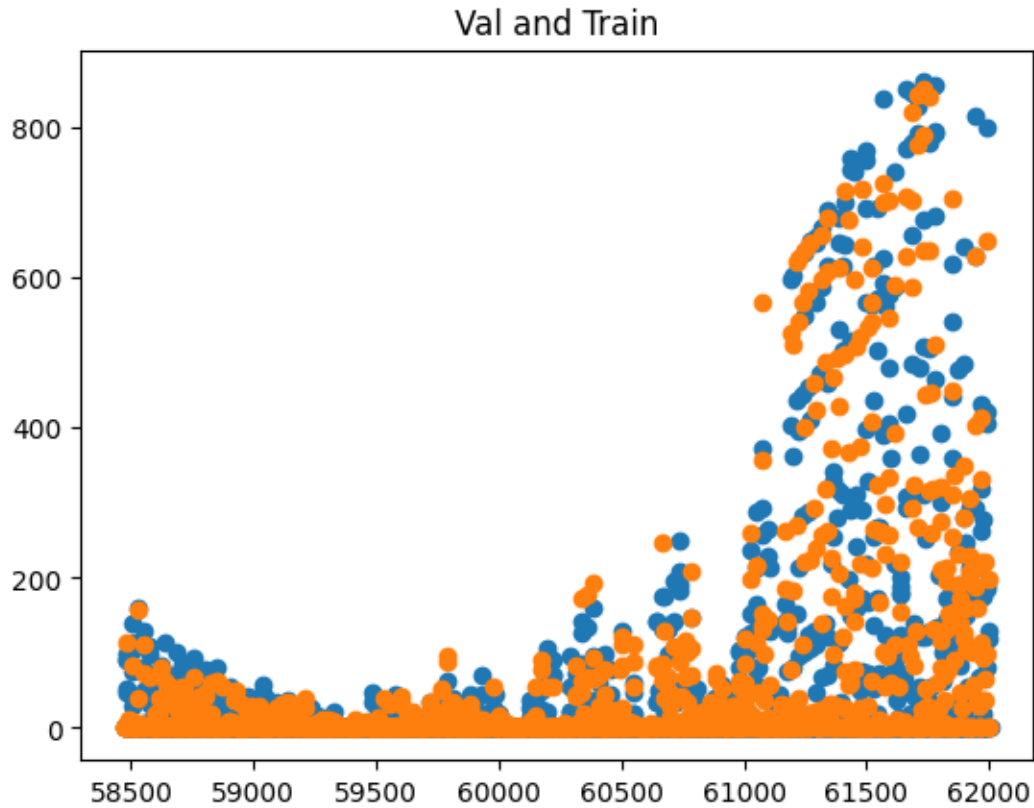
Val and Train

	model	score_test	score_val	pred_time_test	pred_time_val	fit_time	pred_time_test_marginal	pred_time_val_marginal	fit_time_marginal	stack_level	can_infer	fit_order
0	LightGBMXT	-10.416372	-13.714202	0.127264	0.263104	126.649451	0.127264	0.263104	126.649451	1	True	1
1	WeightedEnsemble_L2	-10.771986	-11.964269	0.327276	0.257611	136.459363	0.001567	0.000245	0.067160	2	True	8
2	LightGBM_r118	-11.610208	-12.655171	0.093102	0.147992	43.656658	0.093102	0.147992	43.656658	1	True	5
3	NeuralNetTorch	-12.551883	-13.334396	0.008101	0.010194	20.553850	0.008101	0.010194	20.553850	1	True	4
4	NeuralNetFastAI_r145	-13.092760	-13.791110	0.017274	0.023354	38.659627	0.017274	0.023354	38.659627	1	True	7
5	ExtraTreesMSE	-13.103427	-15.476124	0.147653	0.049357	3.823860	0.147653	0.049357	3.823860			

```
1         True            2
6       NeuralNetFastAI   -14.050613 -14.962268          0.043640          0.019394
17.809215                   0.043640                  0.019394          17.809215
1         True            3
7   NeuralNetFastAI_r51   -14.650156 -14.186801          0.015939          0.007075
11.888993                   0.015939                  0.007075          11.888993
1         True            6
```

Test
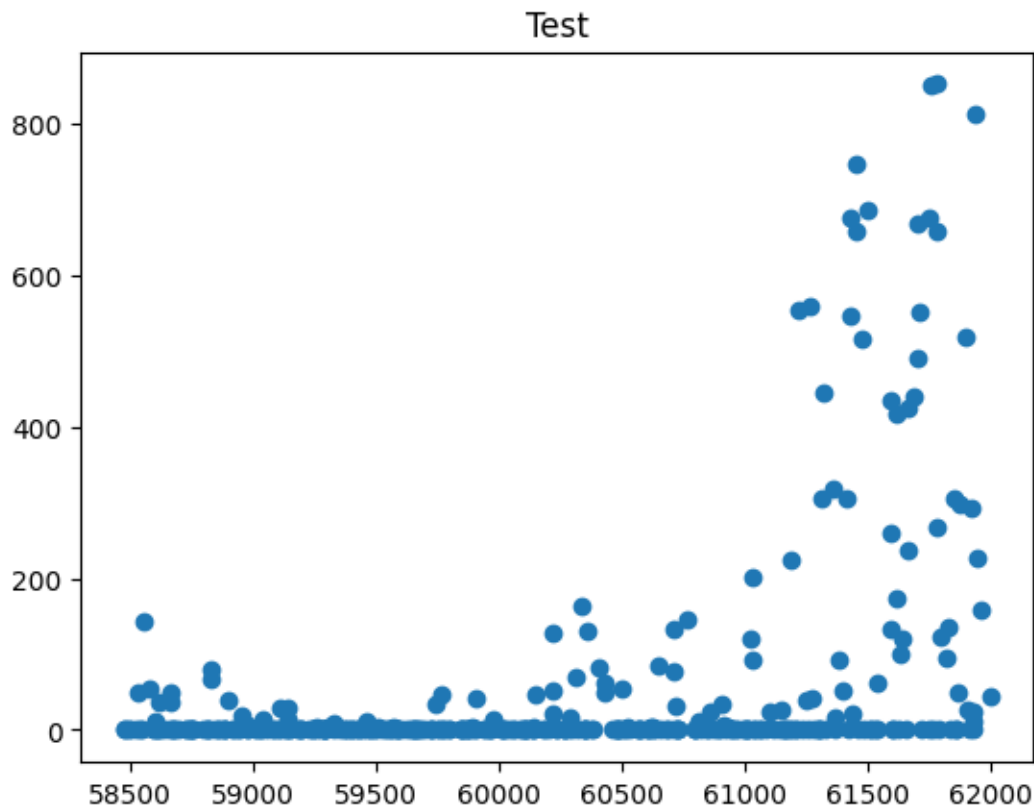

```
[59]: loc = "C"
      predictors[2] = fit_predictor_for_location(loc)
      leaderboards[2] = leaderboard_for_location(2, loc)
```

```
Presets specified: ['experimental_zeroshot_hpo_hybrid']
Stack configuration (auto_stack=True): num_stack_levels=0, num_bag_folds=0,
num_bag_sets=0
Beginning AutoGluon training …
AutoGluon will save models to "AutogluonModels/submission_134_C/"
AutoGluon Version:   0.8.2
Python Version:      3.10.12
Operating System:    Darwin
Platform Machine:    arm64
Platform Version:    Darwin Kernel Version 22.1.0: Sun Oct  9 20:15:09 PDT 2022;
```

```
root:xnu-8792.41.9~2/RELEASE_ARM64_T6000
Disk Space Avail:    100.63 GB / 494.38 GB (20.4%)
Train Data Rows:     24073
Train Data Columns: 47
Tuning Data Rows:     1481
Tuning Data Columns: 47
Label Column: y
Preprocessing data …
AutoGluon infers your prediction problem is: 'regression' (because dtype of
label-column == float and label-values can't be converted to int).
        Label info (max, min, mean, stddev): (999.6, -0.0, 80.87539, 169.67845)
        If 'regression' is not the correct problem_type, please manually specify
the problem_type parameter during predictor init (You may specify problem_type
as one of: ['binary', 'multiclass', 'regression'])
Using Feature Generators to preprocess the data …
Fitting AutoMLPipelineFeatureGenerator…
        Available Memory:                    3110.38 MB
        Train Data (Original)  Memory Usage: 10.89 MB (0.3% of available memory)
        Inferring data type of each feature based on column values. Set
feature_metadata_in to manually specify special dtypes of the features.
        Stage 1 Generators:
                Fitting AsTypeFeatureGenerator…
                        Note: Converting 2 features to boolean dtype as they
only contain 2 unique values.
        Stage 2 Generators:
                Fitting FillNaFeatureGenerator…
        Stage 3 Generators:
                Fitting IdentityFeatureGenerator…
        Stage 4 Generators:
                Fitting DropUniqueFeatureGenerator…
        Stage 5 Generators:
                Fitting DropDuplicatesFeatureGenerator…
        Useless Original Features (Count: 3): ['elevation:m', 'snow_drift:idx',
'location']
                These features carry no predictive signal and should be manually
investigated.
                This is typically a feature which has the same value for all
rows.
                These features do not need to be present at inference time.
        Types of features in original data (raw dtype, special dtypes):
                ('float', []) : 43 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
                ('int', [])   :  1 | ['is_estimated']
        Types of features in processed data (raw dtype, special dtypes):
                ('float', [])    : 42 | ['absolute_humidity_2m:gm3',
'air_density_2m:kgm3', 'ceiling_height_agl:m', 'clear_sky_energy_1h:J',
'clear_sky_rad:W', …]
```

34

```
                    ('int', ['bool']) :  2 | ['snow_density:kgm3', 'is_estimated']
        0.1s = Fit runtime
        44 features in original data used to generate 44 features in processed
data.
        Train Data (Processed) Memory Usage: 8.64 MB (0.3% of available memory)
Data preprocessing and feature engineering runtime = 0.1s …
AutoGluon will gauge predictive performance using evaluation metric:
'mean_absolute_error'
        This metric's sign has been flipped to adhere to being higher_is_better.
The metric score can be multiplied by -1 to get the metric value.
        To change this, specify the eval_metric parameter of Predictor()
Warning: use_bag_holdout=True, but bagged mode is not enabled. use_bag_holdout
will be ignored.
User-specified model hyperparameters to be fit:
{
        'NN_TORCH': {},
        'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}],
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}},
{'extra_trees': True, 'feature_fraction': 0.7832570544199176, 'learning_rate':
0.021720607471727896, 'min_data_in_leaf': 3, 'num_leaves': 21, 'ag_args':
{'name_suffix': '_r118', 'priority': 17}}],
        'FASTAI': [{'bs': 1024, 'emb_drop': 0.6046989241462619, 'epochs': 48,
'layers': [200, 100, 50], 'lr': 0.00775309042164966, 'ps': 0.09244767444160731,
'ag_args': {'name_suffix': '_r51', 'priority': 12}}],
}
Fitting 5 L1 models …
Fitting model: KNeighborsUnif …
        -19.8149         = Validation score   (-mean_absolute_error)
        0.03s    = Training   runtime
        4.73s    = Validation runtime
Fitting model: LightGBMXT …

[1000]  valid_set's l1: 12.5619
[2000]  valid_set's l1: 11.7336
[3000]  valid_set's l1: 11.5889

        -11.5594         = Validation score   (-mean_absolute_error)
        25.45s   = Training   runtime
        0.04s    = Validation runtime
Fitting model: NeuralNetTorch …
        -12.2202         = Validation score   (-mean_absolute_error)
        43.72s   = Training   runtime
        0.01s    = Validation runtime
Fitting model: LightGBM_r118 …

[1000]  valid_set's l1: 13.5054
[2000]  valid_set's l1: 12.4438
[3000]  valid_set's l1: 11.89
[4000]  valid_set's l1: 11.6741
```

```
[5000]  valid_set's l1: 11.4779
[6000]  valid_set's l1: 11.3679
[7000]  valid_set's l1: 11.316
[8000]  valid_set's l1: 11.2841
[9000]  valid_set's l1: 11.2212
[10000] valid_set's l1: 11.195

        -11.1909          = Validation score    (-mean_absolute_error)
        82.2s     = Training    runtime
        0.28s     = Validation runtime
Fitting model: NeuralNetFastAI_r51 …
        -13.4989          = Validation score    (-mean_absolute_error)
        68.52s    = Training    runtime
        0.02s     = Validation runtime
Fitting model: WeightedEnsemble_L2 …
        -10.7391          = Validation score    (-mean_absolute_error)
        0.08s     = Training    runtime
        0.0s      = Validation runtime
AutoGluon training complete, total runtime = 226.87s … Best model:
"WeightedEnsemble_L2"
TabularPredictor saved. To load, use: predictor =
TabularPredictor.load("AutogluonModels/submission_134_C/")
Evaluation: mean_absolute_error on test data: -11.096687568183471
        Note: Scores are always higher_is_better. This metric score can be
multiplied by -1 to get the metric value.
Evaluations on test data:
{
    "mean_absolute_error": -11.096687568183471,
    "root_mean_squared_error": -26.67684023698341,
    "mean_squared_error": -711.6538050295371,
    "r2": 0.9278640745308446,
    "pearsonr": 0.9652418428251913,
    "median_absolute_error": -0.5633931159973145
}

Evaluation on test data:
-11.096687568183471
```
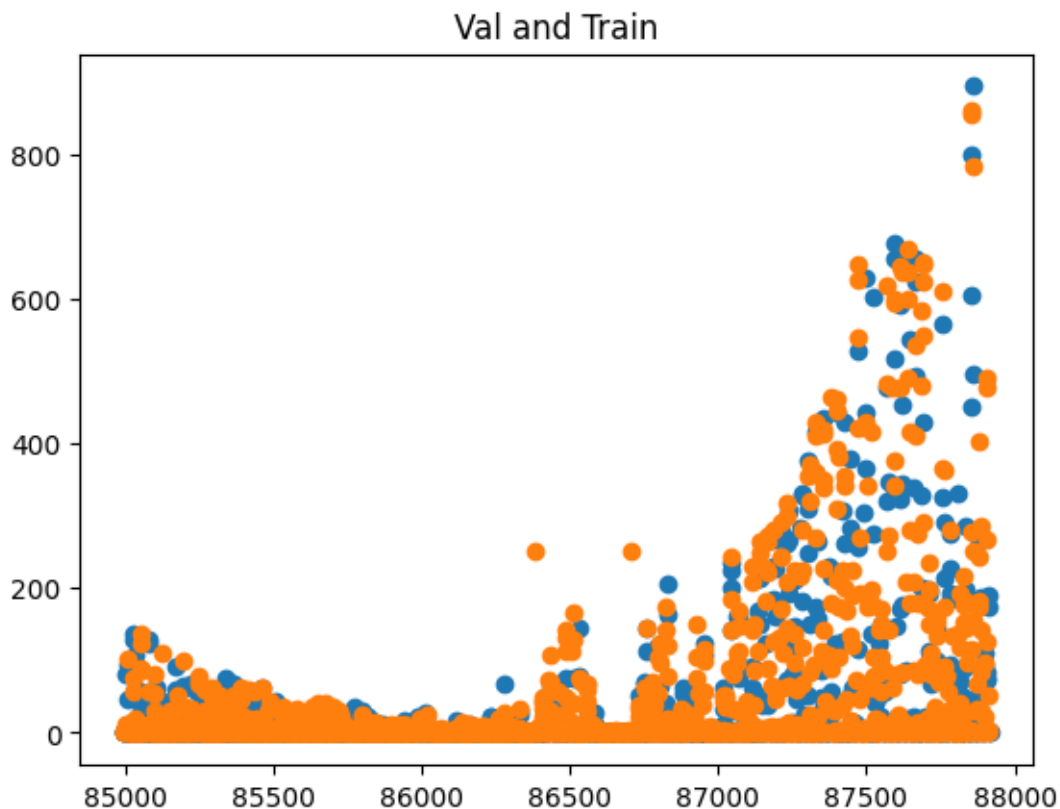
Val and Train

```
            model  score_test  score_val  pred_time_test  pred_time_val
fit_time  pred_time_test_marginal  pred_time_val_marginal  fit_time_marginal
stack_level  can_infer  fit_order
0  WeightedEnsemble_L2  -11.096688 -10.739141      12.911151       5.084761
220.005957            0.002081                0.000204            0.083005
2      True           6
1        LightGBM_r118  -11.531620 -11.190891       0.322478       0.276384
82.199734            0.322478                0.276384           82.199734
1      True           4
2          LightGBMXT  -12.545529 -11.559353       0.133682       0.036625
25.454035            0.133682                0.036625           25.454035
1      True           2
3        NeuralNetTorch  -13.048661 -12.220150       0.092223       0.014784
43.720339            0.092223                0.014784           43.720339
1      True           3
4  NeuralNetFastAI_r51  -14.205124 -13.498880       0.032736       0.023557
68.520166            0.032736                0.023557           68.520166
1      True           5
5        KNeighborsUnif  -20.049167 -19.814903      12.327951       4.733207
0.028678            12.327951                4.733207            0.028678
1      True           1
```

Test
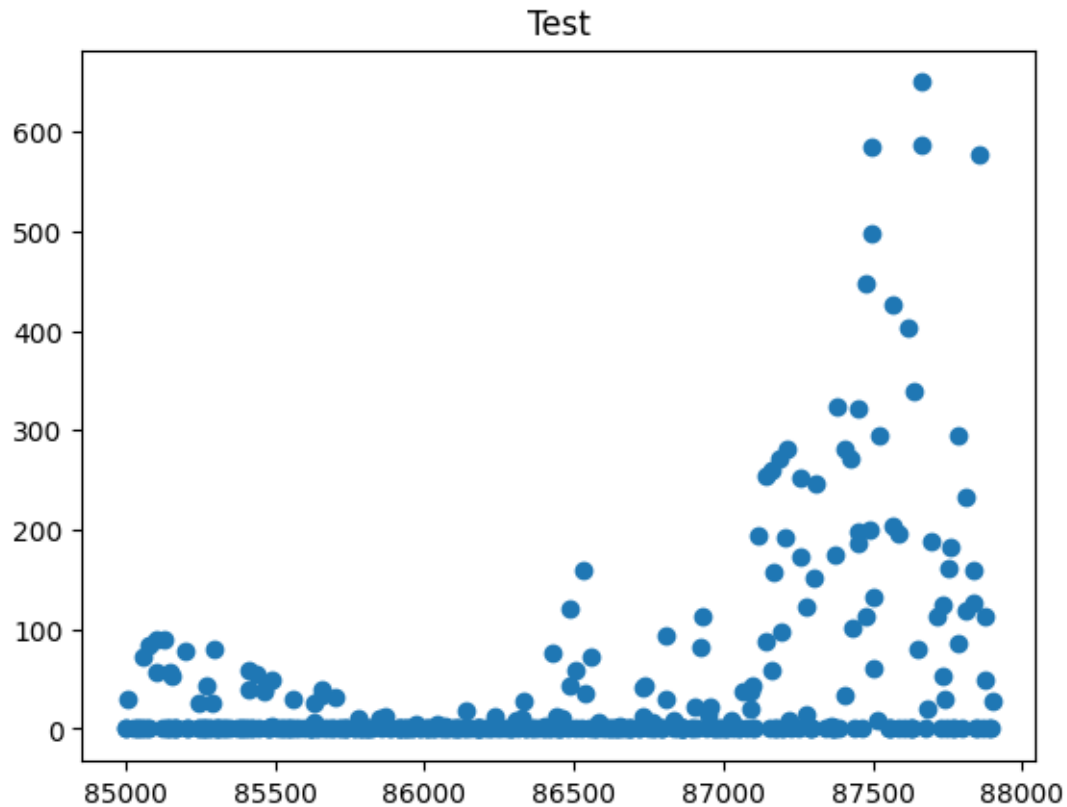
```
[60]: # save leaderboards to csv
      pd.concat(leaderboards).to_csv(f"leaderboards/{new_filename}.csv")


      for i in range(len(predictors)):
          print(f"Predictor {i}:")
          print(predictors[i].
       ↪info()["model_info"]["WeightedEnsemble_L2"]["children_info"]["S1F1"]["model_weights"])
```

```
Predictor 0:
{'LightGBMXT': 0.5147058823529411, 'NeuralNetTorch': 0.4852941176470588}
Predictor 1:
{'ExtraTreesMSE': 0.015151515151515152, 'NeuralNetFastAI': 0.045454545454545456,
'NeuralNetTorch': 0.21212121212121213, 'LightGBM_r118': 0.4393939393939394,
'NeuralNetFastAI_r51': 0.09090909090909091, 'NeuralNetFastAI_r145':
0.19696969696969696}
Predictor 2:
{'KNeighborsUnif': 0.06756756756756757, 'LightGBMXT': 0.13513513513513514,
'NeuralNetTorch': 0.3108108108108108, 'LightGBM_r118': 0.4594594594594595,
'NeuralNetFastAI_r51': 0.02702702702702703}
```

5 Submit

```
[61]: import pandas as pd
      import matplotlib.pyplot as plt

      future_test_data = TabularDataset('X_test_raw.csv')
      future_test_data["ds"] = pd.to_datetime(future_test_data["ds"])
```

Loaded data from: X_test_raw.csv | Columns = 48 / 48 | Rows = 4608 -> 4608

```
[62]: test_ids = TabularDataset('test.csv')
      test_ids["time"] = pd.to_datetime(test_ids["time"])
      # merge test_data with test_ids
      future_test_data_merged = pd.merge(future_test_data, test_ids, how="inner",␣
       ↪right_on=["time", "location"], left_on=["ds", "location"])
```

Loaded data from: test.csv | Columns = 4 / 4 | Rows = 2160 -> 2160

```
[63]: # predict, grouped by location
      predictions = []
      location_map = {
          "A": 0,
          "B": 1,
          "C": 2
      }
      for loc, group in future_test_data.groupby('location'):
          i = location_map[loc]
          subset = future_test_data_merged[future_test_data_merged["location"] ==␣
       ↪loc].reset_index(drop=True)
          pred = predictors[i].predict(subset)
          subset["prediction"] = pred
          predictions.append(subset)

          # get past predictions
          tuning_data.loc[tuning_data["location"] == loc, "prediction"] = ␣
       ↪predictors[i].predict(tuning_data[tuning_data["location"] == loc])
          if use_test_data:
              # get predictions for local test_data
              test_data.loc[test_data["location"] == loc, "prediction"] = ␣
       ↪predictors[i].predict(test_data[test_data["location"] == loc])
```

```
[64]: for loc, idx in location_map.items():
          fig, ax = plt.subplots(figsize=(20, 10))
          # plot train data
          train_data[train_data["location"]==loc].plot(x='ds', y='y', ax=ax,␣
       ↪label="train data")
          tuning_data[tuning_data["location"]==loc].plot(x='ds', y='y', ax=ax,␣
       ↪label="tune data")
```

```
    if use_test_data:
        test_data[test_data["location"]==loc].plot(x='ds', y='y', ax=ax,␣
↪label="test data")

    # plot predictions
    predictions[idx].plot(x='ds', y='prediction', ax=ax, label="predictions")

    # plot past predictions
    tuning_data[tuning_data["location"]==loc].plot(x='ds', y='prediction',␣
↪ax=ax, label="past predictions tune")
    if use_test_data:
        test_data[test_data["location"]==loc].plot(x='ds', y='prediction',␣
↪ax=ax, label="past predictions test")


    ax.set_title(f"Predictions for location {loc}")
```

Predictions for location B

Predictions for location C
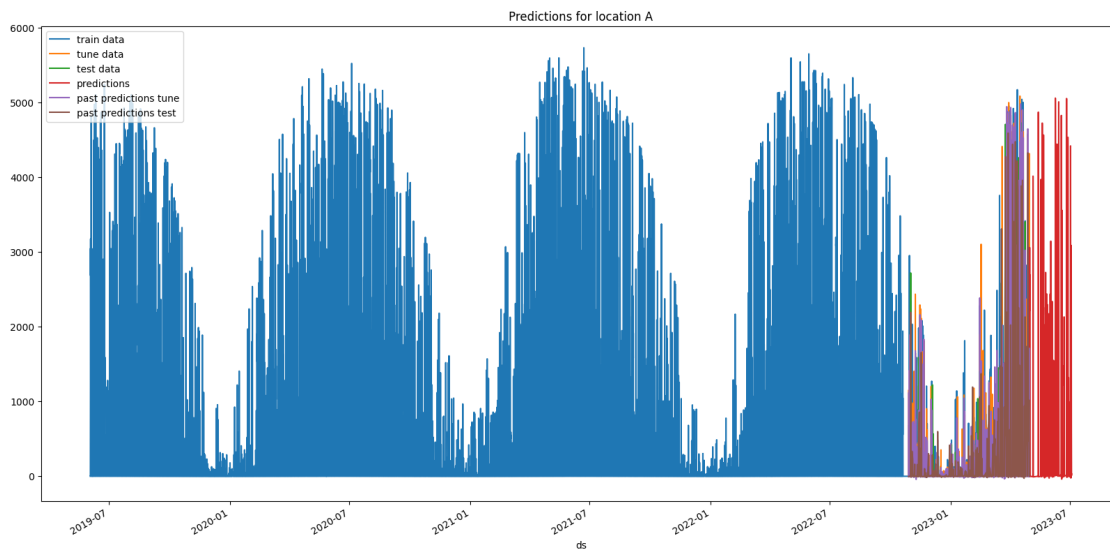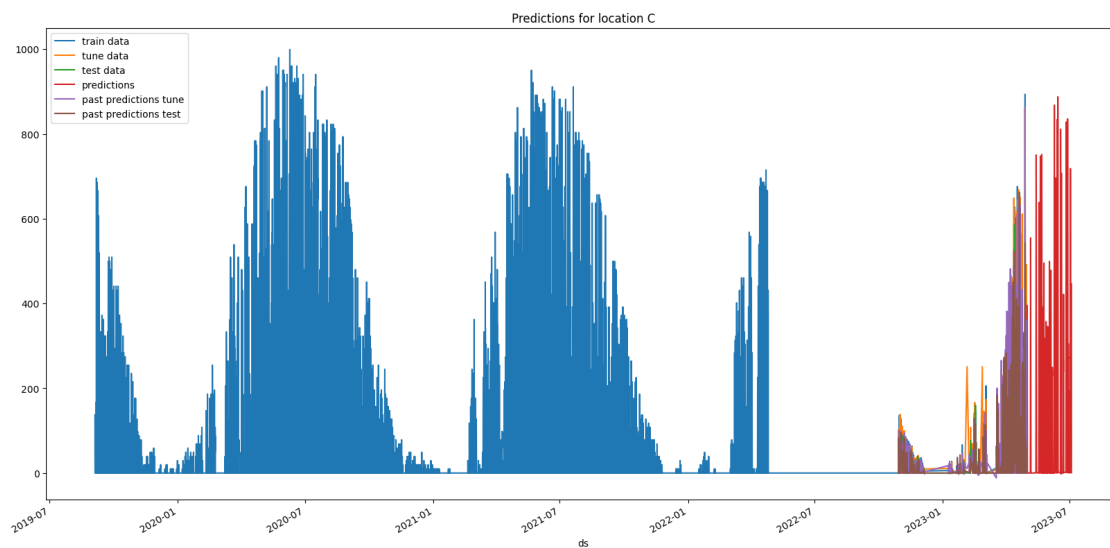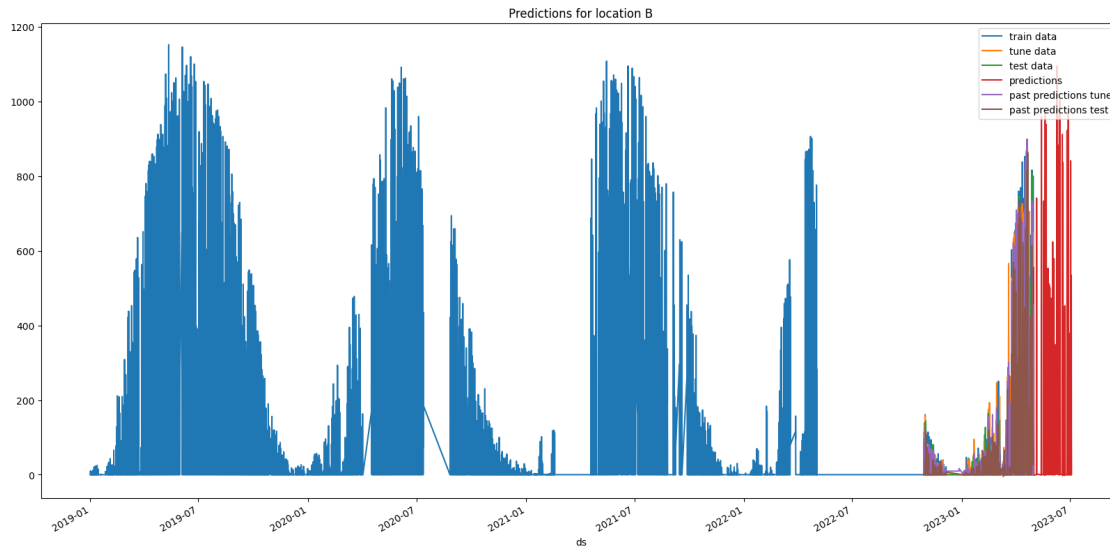
```
[65]: temp_predictions = [prediction.copy() for prediction in predictions]
      if clip_predictions:
          # clip predictions smaller than 0 to 0
          for pred in temp_predictions:
              # print smallest prediction
              print("Smallest prediction:", pred["prediction"].min())
              pred.loc[pred["prediction"] < 0, "prediction"] = 0
              print("Smallest prediction after clipping:", pred["prediction"].min())
```

```
# concatenate predictions
submissions_df = pd.concat(temp_predictions)
submissions_df = submissions_df[["id", "prediction"]]
submissions_df
```

Smallest prediction: -37.823605
Smallest prediction after clipping: 0.0
Smallest prediction: -2.3058283
Smallest prediction after clipping: 0.0
Smallest prediction: -2.1818242
Smallest prediction after clipping: 0.0

[65]:
```
        id   prediction
0        0     0.000000
1        1     0.000000
2        2     0.000000
3        3    16.584423
4        4   304.872498
..      ...         ...
715   2155    63.452984
716   2156    34.039520
717   2157     5.225869
718   2158     0.645353
719   2159     0.920492

[2160 rows x 2 columns]
```

[66]:
```
# Save the submission
print(f"Saving submission to submissions/{new_filename}.csv")
submissions_df.to_csv(os.path.join('submissions', f"{new_filename}.csv"),↵
 ↪index=False)
```

Saving submission to submissions/submission_134.csv

[67]:
```
# feature importance
print("\033[1m" + "Calculating feature importance for location A..." +↵
 ↪"\033[0m")
print(predictors[0].feature_importance(feature_stage="original",↵
 ↪data=test_data[test_data["location"] == "A"], time_limit=60*10))
print("\033[1m" + "Calculating feature importance for location B..." +↵
 ↪"\033[0m")
print(predictors[1].feature_importance(feature_stage="original",↵
 ↪data=test_data[test_data["location"] == "B"], time_limit=60*10))
print("\033[1m" + "Calculating feature importance for location C..." +↵
 ↪"\033[0m")
print(predictors[2].feature_importance(feature_stage="original",↵
 ↪data=test_data[test_data["location"] == "C"], time_limit=60*10))
```

These features in provided data are not utilized by the predictor and will be ignored: ['ds', 'elevation:m', 'snow_drift:idx', 'location', 'prediction']

Calculating feature importance for location A…

Computing feature importance via permutation shuffling for 44 features using 361 rows with 10 shuffle sets… Time limit: 600s…
```
        88.72s  = Expected runtime (8.87s per shuffle set)
        53.07s  = Actual runtime (Completed 10 of 10 shuffle sets)
```
These features in provided data are not utilized by the predictor and will be ignored: ['ds', 'elevation:m', 'location', 'prediction']
Computing feature importance via permutation shuffling for 45 features using 361 rows with 10 shuffle sets… Time limit: 600s…

	importance	stddev	p_value	n
direct_rad:W	7.966799e+01	6.196518	8.202915e-12	10
clear_sky_rad:W	7.633021e+01	12.170163	4.887011e-09	10
diffuse_rad:W	5.799179e+01	10.525372	1.526728e-08	10
clear_sky_energy_1h:J	3.031519e+01	10.381896	3.459234e-06	10
sun_elevation:d	2.939179e+01	7.686730	3.608322e-07	10
sun_azimuth:d	2.394524e+01	12.523500	9.566792e-05	10
diffuse_rad_1h:J	1.427764e+01	4.402150	1.448702e-06	10
direct_rad_1h:J	1.398430e+01	5.747896	1.509840e-05	10
effective_cloud_cover:p	1.397670e+01	4.436740	1.846918e-06	10
total_cloud_cover:p	1.084766e+01	3.430958	1.791842e-06	10
snow_water:kgm2	7.979610e+00	5.356829	5.518101e-04	10
relative_humidity_1000hPa:p	7.824189e+00	2.658841	3.245806e-06	10
ceiling_height_agl:m	5.896515e+00	3.069501	9.241565e-05	10
precip_type_5min:idx	4.407535e+00	4.810146	8.832219e-03	10
cloud_base_agl:m	4.323458e+00	2.144760	6.453748e-05	10
is_day:idx	4.234546e+00	1.755630	1.616799e-05	10
is_in_shadow:idx	4.195547e+00	1.843094	2.545761e-05	10
pressure_100m:hPa	3.815979e+00	2.712117	8.008382e-04	10
wind_speed_10m:ms	3.776024e+00	1.275661	3.092497e-06	10
wind_speed_v_10m:ms	3.342700e+00	2.037866	2.869482e-04	10
visibility:m	3.198419e+00	1.971327	3.094968e-04	10
fresh_snow_6h:cm	3.072502e+00	2.323215	1.184209e-03	10
wind_speed_u_10m:ms	2.988759e+00	4.273209	2.714485e-02	10
pressure_50m:hPa	2.501973e+00	1.564861	3.423738e-04	10
snow_depth:cm	2.467273e+00	1.609154	4.550271e-04	10
msl_pressure:hPa	2.072250e+00	2.394007	1.147504e-02	10
fresh_snow_24h:cm	2.062518e+00	4.520783	9.149065e-02	10
air_density_2m:kgm3	1.980115e+00	1.604045	1.799604e-03	10
sfc_pressure:hPa	1.837239e+00	1.883139	6.515049e-03	10
precip_5min:mm	1.833741e+00	2.442299	2.080664e-02	10
super_cooled_liquid_water:kgm2	1.543328e+00	1.515049	5.233747e-03	10
fresh_snow_1h:cm	1.070496e+00	1.074131	5.853889e-03	10
t_1000hPa:K	9.737481e-01	1.600851	4.328374e-02	10
snow_melt_10min:mm	8.978923e-01	1.390949	3.580554e-02	10

```
fresh_snow_3h:cm              7.902873e-01   1.406984   5.471397e-02   10
absolute_humidity_2m:gm3      6.890160e-01   0.912385   2.034222e-02   10
rain_water:kgm2               5.316596e-01   0.524287   5.357342e-03   10
fresh_snow_12h:cm             5.251203e-01   3.366204   3.168039e-01   10
dew_point_2m:K                3.666395e-01   1.146486   1.691486e-01   10
prob_rime:p                   3.051479e-01   0.548112   5.608503e-02   10
snow_density:kgm3             2.111609e-01   1.576556   3.409184e-01   10
is_estimated                  4.490987e-08   0.000000   5.000000e-01   10
wind_speed_w_1000hPa:ms       4.226811e-08   0.000000   5.000000e-01   10
dew_or_rime:idx              -5.206986e-01   1.231628   8.929724e-01   10


                                 p99_high        p99_low
direct_rad:W                  8.603607e+01   7.329990e+01
clear_sky_rad:W               8.883734e+01   6.382308e+01
diffuse_rad:W                 6.880859e+01   4.717499e+01
clear_sky_energy_1h:J         4.098454e+01   1.964583e+01
sun_elevation:d               3.729135e+01   2.149223e+01
sun_azimuth:d                 3.681550e+01   1.107499e+01
diffuse_rad_1h:J              1.880168e+01   9.753601e+00
direct_rad_1h:J               1.989135e+01   8.077258e+00
effective_cloud_cover:p       1.853628e+01   9.417111e+00
total_cloud_cover:p           1.437362e+01   7.321705e+00
snow_water:kgm2               1.348476e+01   2.474460e+00
relative_humidity_1000hPa:p   1.055665e+01   5.091729e+00
ceiling_height_agl:m          9.051005e+00   2.742025e+00
precip_type_5min:idx          9.350865e+00  -5.357949e-01
cloud_base_agl:m              6.527603e+00   2.119313e+00
is_day:idx                    6.038786e+00   2.430306e+00
is_in_shadow:idx              6.089673e+00   2.301421e+00
pressure_100m:hPa             6.603190e+00   1.028768e+00
wind_speed_10m:ms             5.087007e+00   2.465042e+00
wind_speed_v_10m:ms           5.436991e+00   1.248409e+00
visibility:m                  5.224328e+00   1.172509e+00
fresh_snow_6h:cm              5.460043e+00   6.849609e-01
wind_speed_u_10m:ms           7.380285e+00  -1.402768e+00
pressure_50m:hPa              4.110162e+00   8.937838e-01
snow_depth:cm                 4.120981e+00   8.135643e-01
msl_pressure:hPa              4.532542e+00  -3.880429e-01
fresh_snow_24h:cm             6.708473e+00  -2.583437e+00
air_density_2m:kgm3           3.628574e+00   3.316571e-01
sfc_pressure:hPa              3.772519e+00  -9.804137e-02
precip_5min:mm                4.343663e+00  -6.761803e-01
super_cooled_liquid_water:kgm2 3.100326e+00  -1.366959e-02
fresh_snow_1h:cm              2.174368e+00  -3.337505e-02
t_1000hPa:K                   2.618923e+00  -6.714272e-01
snow_melt_10min:mm            2.327354e+00  -5.315699e-01
fresh_snow_3h:cm              2.236228e+00  -6.556536e-01
absolute_humidity_2m:gm3      1.626664e+00  -2.486316e-01
```

```
rain_water:kgm2                  1.070463e+00 -7.143807e-03
fresh_snow_12h:cm                3.984528e+00 -2.934288e+00
dew_point_2m:K                   1.544870e+00 -8.115905e-01
prob_rime:p                      8.684358e-01 -2.581400e-01
snow_density:kgm3                1.831368e+00 -1.409047e+00
is_estimated                     4.490987e-08  4.490987e-08
wind_speed_w_1000hPa:ms          4.226811e-08  4.226811e-08
dew_or_rime:idx                  7.450309e-01 -1.786428e+00
```
Calculating feature importance for location B…

```
        307.44s = Expected runtime (30.74s per shuffle set)
        55.82s  = Actual runtime (Completed 10 of 10 shuffle sets)
```
These features in provided data are not utilized by the predictor and will be
ignored: ['ds', 'elevation:m', 'snow_drift:idx', 'location', 'prediction']
Computing feature importance via permutation shuffling for 44 features using 360
rows with 10 shuffle sets… Time limit: 600s…

	importance	stddev	p_value	n \
clear_sky_rad:W	3.089353e+01	2.862590	3.928157e-11	10
direct_rad:W	1.758631e+01	1.326535	6.232521e-12	10
diffuse_rad:W	1.292878e+01	1.449306	2.149220e-10	10
sun_elevation:d	8.974525e+00	1.129608	6.024637e-10	10
sun_azimuth:d	6.951743e+00	1.180075	8.485674e-09	10
clear_sky_energy_1h:J	6.795127e+00	1.047551	3.631051e-09	10
effective_cloud_cover:p	3.173118e+00	0.654582	4.670364e-08	10
direct_rad_1h:J	3.120853e+00	0.435006	1.489493e-09	10
diffuse_rad_1h:J	3.112522e+00	0.729352	1.407778e-07	10
relative_humidity_1000hPa:p	1.751303e+00	0.483333	5.701084e-07	10
is_in_shadow:idx	1.597443e+00	0.199220	5.550254e-10	10
fresh_snow_24h:cm	1.452799e+00	0.497733	3.470695e-06	10
snow_water:kgm2	1.280799e+00	0.406230	1.833939e-06	10
wind_speed_v_10m:ms	1.147136e+00	0.256486	9.390015e-08	10
sfc_pressure:hPa	1.121787e+00	0.288322	3.110949e-07	10
visibility:m	1.118796e+00	0.305319	5.182709e-07	10
cloud_base_agl:m	1.087815e+00	0.273498	2.576902e-07	10
total_cloud_cover:p	9.360905e-01	0.258885	5.802443e-07	10
pressure_100m:hPa	9.270775e-01	0.325786	4.273716e-06	10
pressure_50m:hPa	8.658717e-01	0.314427	5.585112e-06	10
msl_pressure:hPa	8.484935e-01	0.205324	1.856428e-07	10
ceiling_height_agl:m	6.605458e-01	0.324262	5.964112e-05	10
wind_speed_10m:ms	6.583129e-01	0.189969	8.310688e-07	10
t_1000hPa:K	6.118650e-01	0.236174	9.147006e-06	10
snow_density:kgm3	5.429378e-01	0.505616	3.963566e-03	10
is_day:idx	4.907764e-01	0.092593	2.143246e-08	10
air_density_2m:kgm3	4.703118e-01	0.357817	1.230291e-03	10
wind_speed_u_10m:ms	3.440242e-01	0.276338	1.711148e-03	10
super_cooled_liquid_water:kgm2	3.356328e-01	0.321861	4.632723e-03	10
precip_type_5min:idx	2.899514e-01	0.306324	7.560363e-03	10
dew_point_2m:K	2.881253e-01	0.123478	2.098102e-05	10

```
fresh_snow_12h:cm              2.772887e-01  0.315221  1.066996e-02  10
fresh_snow_6h:cm               2.697109e-01  0.227036  2.254014e-03  10
precip_5min:mm                 2.373693e-01  0.313540  2.014490e-02  10
rain_water:kgm2                2.218524e-01  0.180006  1.816843e-03  10
snow_depth:cm                  1.467803e-01  0.140861  4.650676e-03  10
absolute_humidity_2m:gm3       1.399695e-01  0.065372  4.084992e-05  10
prob_rime:p                    1.062888e-01  0.199169  6.288190e-02  10
fresh_snow_1h:cm               1.057443e-01  0.099328  4.151120e-03  10
dew_or_rime:idx                7.235418e-02  0.080144  9.468436e-03  10
fresh_snow_3h:cm               7.209251e-02  0.087237  1.405900e-02  10
snow_melt_10min:mm             3.663396e-02  0.228162  3.119229e-01  10
is_estimated                   0.000000e+00  0.000000  5.000000e-01  10
snow_drift:idx                -5.283516e-09  0.000000  5.000000e-01  10
wind_speed_w_1000hPa:ms       -5.283516e-09  0.000000  5.000000e-01  10


                                 p99_high       p99_low
clear_sky_rad:W                3.383538e+01  2.795168e+01
direct_rad:W                   1.894957e+01  1.622304e+01
diffuse_rad:W                  1.441822e+01  1.143935e+01
sun_elevation:d                1.013541e+01  7.813640e+00
sun_azimuth:d                  8.164492e+00  5.738994e+00
clear_sky_energy_1h:J          7.871682e+00  5.718571e+00
effective_cloud_cover:p        3.845824e+00  2.500411e+00
direct_rad_1h:J                3.567903e+00  2.673802e+00
diffuse_rad_1h:J               3.862068e+00  2.362976e+00
relative_humidity_1000hPa:p    2.248019e+00  1.254587e+00
is_in_shadow:idx               1.802180e+00  1.392707e+00
fresh_snow_24h:cm              1.964313e+00  9.412840e-01
snow_water:kgm2                1.698277e+00  8.633209e-01
wind_speed_v_10m:ms            1.410724e+00  8.835485e-01
sfc_pressure:hPa               1.418092e+00  8.254824e-01
visibility:m                   1.432569e+00  8.050229e-01
cloud_base_agl:m               1.368886e+00  8.067434e-01
total_cloud_cover:p            1.202143e+00  6.700376e-01
pressure_100m:hPa              1.261883e+00  5.922715e-01
pressure_50m:hPa               1.189005e+00  5.427385e-01
msl_pressure:hPa               1.059502e+00  6.374845e-01
ceiling_height_agl:m           9.937865e-01  3.273051e-01
wind_speed_10m:ms              8.535422e-01  4.630837e-01
t_1000hPa:K                    8.545787e-01  3.691512e-01
snow_density:kgm3              1.062553e+00  2.332255e-02
is_day:idx                     5.859336e-01  3.956193e-01
air_density_2m:kgm3            8.380361e-01  1.025875e-01
wind_speed_u_10m:ms            6.280135e-01  6.003497e-02
super_cooled_liquid_water:kgm2 6.664051e-01  4.860577e-03
precip_type_5min:idx           6.047566e-01 -2.485390e-02
dew_point_2m:K                 4.150219e-01  1.612288e-01
fresh_snow_12h:cm              6.012377e-01 -4.666037e-02
```

```
fresh_snow_6h:cm                   5.030330e-01   3.638877e-02
precip_5min:mm                     5.595911e-01  -8.485250e-02
rain_water:kgm2                    4.068427e-01   3.686212e-02
snow_depth:cm                      2.915413e-01   2.019220e-03
absolute_humidity_2m:gm3           2.071515e-01   7.278755e-02
prob_rime:p                        3.109729e-01  -9.839529e-02
fresh_snow_1h:cm                   2.078223e-01   3.666258e-03
dew_or_rime:idx                    1.547172e-01  -1.000885e-02
fresh_snow_3h:cm                   1.617449e-01  -1.755983e-02
snow_melt_10min:mm                 2.711138e-01  -1.978459e-01
is_estimated                       0.000000e+00   0.000000e+00
snow_drift:idx                    -5.283516e-09  -5.283516e-09
wind_speed_w_1000hPa:ms           -5.283516e-09  -5.283516e-09
Calculating feature importance for location C…

        3929.37s              = Expected runtime (392.94s per shuffle set)
```

```
        --------------------------------------------------------------------------------
KeyboardInterrupt                               Traceback (most recent call last)
/Users/jorgensandhaug/Desktop/tdt4173/TDT4173/short_1.ipynb Cell 31 line 7
      <a href='vscode-notebook-cell:/Users/jorgensandhaug/Desktop/tdt4173/
  ↪TDT4173/short_1.ipynb#X43sZmlsZQ%3D%3D?line=4'>5</a> print(predictors[1].
  ↪feature_importance(feature_stage="original",␣
  ↪data=test_data[test_data["location"] == "B"], time_limit=60*10))
      <a href='vscode-notebook-cell:/Users/jorgensandhaug/Desktop/tdt4173/
  ↪TDT4173/short_1.ipynb#X43sZmlsZQ%3D%3D?line=5'>6</a> print("\033[1m" +␣
  ↪"Calculating feature importance for location C…" + "\033[0m")
----> <a href='vscode-notebook-cell:/Users/jorgensandhaug/Desktop/tdt4173/
  ↪TDT4173/short_1.ipynb#X43sZmlsZQ%3D%3D?line=6'>7</a> print(predictors[2].
  ↪feature_importance(feature_stage="original",␣
  ↪data=test_data[test_data["location"] == "C"], time_limit=60*10))

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/
  ↪tabular/predictor/predictor.py:2418, in TabularPredictor.
  ↪feature_importance(self, data, model, features, feature_stage, subsample_size␣
  ↪time_limit, num_shuffle_sets, include_confidence_band, confidence_level,␣
  ↪silent)
   2415 if num_shuffle_sets is None:
   2416     num_shuffle_sets = 10 if time_limit else 5
-> 2418 fi_df = self._learner.get_feature_importance(
   2419     model=model,
   2420     X=data,
   2421     features=features,
   2422     feature_stage=feature_stage,
   2423     subsample_size=subsample_size,
   2424     time_limit=time_limit,
   2425     num_shuffle_sets=num_shuffle_sets,
   2426     silent=silent,
   2427 )
   2429 if include_confidence_band:
```

```
2430        if confidence_level <= 0.5 or confidence_level >= 1.0:

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/
  ↪tabular/learner/abstract_learner.py:870, in AbstractTabularLearner.
  ↪get_feature_importance(self, model, X, y, features, feature_stage,␣
  ↪subsample_size, silent, **kwargs)
    867            X = X.drop(columns=unused_features)
    869        if feature_stage == "original":
--> 870            return trainer._get_feature_importance_raw(
    871                model=model, X=X, y=y, features=features,␣
  ↪subsample_size=subsample_size, transform_func=self.transform_features,␣
  ↪silent=silent, **kwargs
    872            )
    873        X = self.transform_features(X)
    874 else:

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:2622, in AbstractTrainer.
  ↪_get_feature_importance_raw(self, X, y, model, eval_metric, **kwargs)
    2620 model: AbstractModel = self.load_model(model)
    2621 predict_func_kwargs = dict(model=model)
-> 2622 return compute_permutation_feature_importance(
    2623     X=X,
    2624     y=y,
    2625     predict_func=predict_func,
    2626     predict_func_kwargs=predict_func_kwargs,
    2627     eval_metric=eval_metric,
    2628     quantile_levels=self.quantile_levels,
    2629     **kwargs,
    2630 )

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪utils/utils.py:882, in compute_permutation_feature_importance(X, y,␣
  ↪predict_func, eval_metric, features, subsample_size, num_shuffle_sets,␣
  ↪predict_func_kwargs, transform_func, transform_func_kwargs, time_limit,␣
  ↪silent, log_prefix, importance_as_list, random_state, **kwargs)
    880 else:
    881     X_raw_transformed = X_raw if transform_func is None else␣
  ↪transform_func(X_raw, **transform_func_kwargs)
--> 882 y_pred = predict_func(X_raw_transformed, **predict_func_kwargs)
    884 row_index = 0
    885 for feature in parallel_computed_features:

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:737, in AbstractTrainer.predict(self, X, model)
    735     model = self._get_best()
    736 cascade = isinstance(model, list)
--> 737 return self._predict_model(X, model, cascade=cascade)
```

```
File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:2436, in AbstractTrainer._predict_model(self, X,␣
  ↪model, model_pred_proba_dict, cascade)
    2435 def _predict_model(self, X, model, model_pred_proba_dict=None,␣
  ↪cascade=False):
-> 2436     y_pred_proba = self._predict_proba_model(X=X, model=model,␣
  ↪model_pred_proba_dict=model_pred_proba_dict, cascade=cascade)
    2437     return get_pred_from_proba(y_pred_proba=y_pred_proba,␣
  ↪problem_type=self.problem_type)

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:2440, in AbstractTrainer.
  ↪_predict_proba_model(self, X, model, model_pred_proba_dict, cascade)
    2439 def _predict_proba_model(self, X, model, model_pred_proba_dict=None,␣
  ↪cascade=False):
-> 2440     return self.get_pred_proba_from_model(model=model, X=X,␣
  ↪model_pred_proba_dict=model_pred_proba_dict, cascade=cascade)

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:757, in AbstractTrainer.
  ↪get_pred_proba_from_model(self, model, X, model_pred_proba_dict, cascade)
    755 else:
    756     models = [model]
--> 757 model_pred_proba_dict = self.get_model_pred_proba_dict(X=X,␣
  ↪models=models, model_pred_proba_dict=model_pred_proba_dict, cascade=cascade)
    758 if not isinstance(model, str):
    759     model = model.name

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪trainer/abstract_trainer.py:1008, in AbstractTrainer.
  ↪get_model_pred_proba_dict(self, X, models, model_pred_proba_dict,␣
  ↪model_pred_time_dict, record_pred_time, use_val_cache, cascade,␣
  ↪cascade_threshold)
    1006     model_pred_proba_dict[model_name] = model.predict_proba(X,␣
  ↪**preprocess_kwargs)
    1007 else:
-> 1008     model_pred_proba_dict[model_name] = model.predict_proba(X)
    1010 if record_pred_time:
    1011     time_end = time.time()

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
  ↪models/abstract/abstract_model.py:940, in AbstractModel.predict_proba(self, X,␣
  ↪normalize, **kwargs)
    938 if normalize is None:
    939     normalize = self.normalize_pred_probas
--> 940 y_pred_proba = self._predict_proba(X=X, **kwargs)
    941 if normalize:
    942     y_pred_proba = normalize_pred_probas(y_pred_proba, self.problem_type)
```

```
File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/autogluon/core/
 ↪models/abstract/abstract_model.py:955, in AbstractModel._predict_proba(self,␣
 ↪X, **kwargs)
    952 X = self.preprocess(X, **kwargs)
    954 if self.problem_type in [REGRESSION, QUANTILE]:
--> 955     y_pred = self.model.predict(X)
    956     return y_pred
    958 y_pred_proba = self.model.predict_proba(X)


File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/sklearn/
 ↪neighbors/_regression.py:236, in KNeighborsRegressor.predict(self, X)
    220 """Predict the target for the provided data.
    221
    222 Parameters
    (…)
    231     Target values.
    232 """
    233 if self.weights == "uniform":
    234     # In that case, we do not need the distances to perform
    235     # the weighting so we do not compute them.
--> 236     neigh_ind = self.kneighbors(X, return_distance=False)
    237     neigh_dist = None
    238 else:


File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/sklearn/
 ↪neighbors/_base.py:824, in KNeighborsMixin.kneighbors(self, X, n_neighbors,␣
 ↪return_distance)
    817 use_pairwise_distances_reductions = (
    818     self._fit_method == "brute"
    819     and ArgKmin.is_usable_for(
    820         X if X is not None else self._fit_X, self._fit_X, self.
 ↪effective_metric_
    821     )
    822 )
    823 if use_pairwise_distances_reductions:
--> 824     results = ArgKmin.compute(
    825         X=X,
    826         Y=self._fit_X,
    827         k=n_neighbors,
    828         metric=self.effective_metric_,
    829         metric_kwargs=self.effective_metric_params_,
    830         strategy="auto",
    831         return_distance=return_distance,
    832     )
    834 elif (
    835     self._fit_method == "brute" and self.metric == "precomputed" and␣
 ↪issparse(X)
    836 ):
```

```
        837     results = _kneighbors_from_graph(
        838         X, n_neighbors=n_neighbors, return_distance=return_distance
        839     )

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/sklearn/
 ↪metrics/_pairwise_distances_reduction/_dispatcher.py:289, in ArgKmin.
 ↪compute(cls, X, Y, k, metric, chunk_size, metric_kwargs, strategy,␣
 ↪return_distance)
        277     return ArgKmin64.compute(
        278         X=X,
        279         Y=Y,
    (…)
        285         return_distance=return_distance,
        286     )
        288 if X.dtype == Y.dtype == np.float32:
--> 289     return ArgKmin32.compute(
        290         X=X,
        291         Y=Y,
        292         k=k,
        293         metric=metric,
        294         chunk_size=chunk_size,
        295         metric_kwargs=metric_kwargs,
        296         strategy=strategy,
        297         return_distance=return_distance,
        298     )
        300 raise ValueError(
        301     "Only float64 or float32 datasets pairs are supported at this time, "
        302     f"got: X.dtype={X.dtype} and Y.dtype={Y.dtype}."
        303 )

File sklearn/metrics/_pairwise_distances_reduction/_argkmin.pyx:584, in sklearn
 ↪metrics._pairwise_distances_reduction._argkmin.ArgKmin32.compute()

File /opt/homebrew/anaconda3/envs/ag/lib/python3.10/site-packages/threadpoolctl
 ↪py:440, in _ThreadpoolLimiter.__exit__(self, type, value, traceback)
        437 def __enter__(self):
        438     return self
--> 440 def __exit__(self, type, value, traceback):
        441     self.restore_original_limits()
        443 @classmethod
        444 def wrap(cls, controller, *, limits=None, user_api=None):

KeyboardInterrupt:
```

```
[3]:  # save this notebook to submissions folder
      import subprocess
      import os
```

```
subprocess.run(["jupyter", "nbconvert", "--to", "pdf", "--output", os.path.
 ↪join('notebook_pdfs', "submission_134.pdf"), "short_1.ipynb"])
```

```
  Cell In[3], line 4
    subprocess.run(["jupyter", "nbconvert", "--to", "pdf", "--output", os.path.
 ↪join('notebook_pdfs', f"{"submission_134"}.pdf"), "short_1.ipynb"])

 ↪                              ^
SyntaxError: f-string: expecting '}'
```